



Vibrations de plaques multi-exciteurs de grandes dimensions pour la création d'environnements virtuels audio-visuels: approches acoustique, mécanique et perceptive.

Marc Rébillat

► To cite this version:

Marc Rébillat. Vibrations de plaques multi-exciteurs de grandes dimensions pour la création d'environnements virtuels audio-visuels: approches acoustique, mécanique et perceptive.. Acoustique [physics.class-ph]. Ecole Polytechnique X, 2011. Français. NNT: . pastel-00657634

HAL Id: pastel-00657634

<https://pastel.archives-ouvertes.fr/pastel-00657634>

Submitted on 7 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE L'ÉCOLE POLYTECHNIQUE

Spécialité : Mécanique

présentée par

Marc RÉBILLAT

Sujet :

***Vibrations de plaques multi-exciteurs de grandes dimensions
pour la création d'environnements virtuels audio-visuels***
Approches acoustique, mécanique et perceptive

préparée

au LABORATOIRE DE MÉCANIQUE DES SOLIDES (CNRS UMR 7649)

et au LABORATOIRE D'INFORMATIQUE POUR LA MÉCANIQUE ET LES SCIENCES DE L'INGÉNIEUR
(CNRS UPR 3251)

Soutenue le Jeudi 17 Novembre 2011 devant le jury composé de

Daniel MESTRE	Directeur de recherche au CNRS, ISM, Université Aix-Marseille II	Président
Jean-Louis GUYADER	Professeur, INSA Lyon	Rapporteur
Rozenn NICOL	Ingénieur de recherche, Orange Labs	Rapporteur
Jens BLAUERT	Professeur Émérite, Ruhr-Universität Bochum	Examineur
Vincent PAGNEUX	Directeur de recherche au CNRS, LAUM, Université du Maine	Examineur
Étienne CORTEEL	Directeur scientifique (CSO), <i>sonic emotion labs</i>	Encadrant
Brian F.G. KATZ	Chargé de recherche au CNRS, LIMSI, Université Paris-Sud 11	Encadrant
Xavier BOUTILLON	Directeur de recherche au CNRS, LMS, École Polytechnique	Directeur de thèse



FIG. 1 – Affiche de 1902 faisant l'éloge d'un des premiers système de cinéma sonore développé par *Gaumont* (d'après Wikipédia).

*À Jean-Claude, le chercheur de la famille,
qui m'a sans le savoir amené jusque là.*

Remerciements

En premier lieu, je souhaiterais remercier mes encadrants, Étienne et Brian, et plus particulièrement mon directeur, Xavier, pour la confiance qu'ils ont placée en moi en me voyant débarquer début 2008 à la recherche d'un stage de Master 2. C'est à la suite de quelques rendez-vous que nous nous sommes lancés, à partir de rien et sans nous connaître vraiment, dans ce projet de stage et de thèse aux allures farfelues. Merci Étienne, en plus des échanges fructueux que nous avons pu avoir durant ces trois ans, pour la liberté que tu m'as laissée tout au long de la thèse. Même si je ne suis pas forcément allé jusque là où tu l'aurais souhaité, je suis certain que mon successeur s'en chargera. Merci Brian pour ta disponibilité et pour les moyens que tu as mis à ma disposition tout au long de la thèse et qui m'auront permis d'aboutir à un dispositif fonctionnel. Enfin, merci Xavier pour ton ouverture d'esprit, ta curiosité, la rigueur de ta démarche scientifique, et surtout pour m'avoir dit assez souvent : *“Je ne sais pas, essaie quand même et on verra bien.”*

Merci aussi aux différentes personnes qui ont accepté de faire partie du jury de ma thèse. En particulier, merci à Rozenn Nicol et à Jean-Louis Guyader, les rapporteurs, à qui incombait la lourde tâche de lire ce manuscrit assez long et d'en tirer un rapport concis. Merci aussi à Daniel Mestre pour avoir présidé le jury, ainsi qu'à Jens Blauert et à Vincent Pagneux pour m'avoir fait l'honneur d'accepter notre invitation.

Je tiens ensuite à souligner le fait que ce travail de thèse n'aurait jamais pu voir le jour si j'avais travaillé seul. En effet, les différentes rencontres que j'ai pu faire durant cette thèse m'ont permis de combler mes nombreuses incompétences. Merci donc à Kerem, mon mentor en plaques, pour la plongée dans le grand bain en mécanique. Merci à Matthieu Courgeon pour m'avoir entrouvert la lourde porte du monde des graphistes et des informaticiens fous (voir Fig. 2.12). Sans toi, le SMART-I² ne serait pas audio-visuel. Merci à Romain pour ses compétences mathématiques *rock'n roll* qui ont permis l'analyse des non-linéarités. Merci à Gaétan et Tifanie pour m'avoir appris tout ce que je sais en MAX/MSP¹. Merci à Baptiste et Sarah pour leurs *gestes* dans le projet GAVIP qui, il est certain, sera installé un jour au centre Pompidou. Enfin, si le SMART-I² tient debout, c'est aussi grâce à Pascal, Daniel, Yves, Vincent G. et aux outils de l'atelier.

Les compétences sont une chose, mais rien n'est possible sans une ambiance détendue et propice au travail :

- Au LIMSI, *spéciale kassedédi* à Gaétan, *a.k.a.* GLP, qui aura supporté dans son bureau durant trois ans ma présence et ma musique de banlieusard tout en méditant sur tout et n'importe quoi. Plus personne ne me fait la blague du volume à fond à présent... Merci aussi à Tifanie pour les pauses passées dans son bureau en sa présence, mais sans son consentement, parce qu'il y'avait toujours des bonbons. Merci à David qui m'a fait courir, mais qui a triché en coupant jalousement ses *dreadlocks* pour aller plus

1. Promis ! Un jour j'utiliserai des *triggers*.

vite parce que je le dépassais. Merci à Lionel pour les instants musicaux tribaux. Merci à Polo pour n’être jamais allé trop vite avec moi. Merci à Sleub pour être arrivé avant moi et parti après moi. Merci à Albert pour les statistiques. Merci à Christophe pour ce duo inoubliable et improbable *clavicorde & xaphoon* bientôt en boucle sur NRJ. Merci à nos voisins du bâtiment 502 bis pour relativiser nos nuisances sonores. Merci à Nédé pour le café. Et de manière générale, merci à tout les membres du LMSI pour leur générosité.

- Au LMS, un merci tout particulier aux membres de la *Fanfare à Gérard* : Éva, Matthieu B., Raman, Anders, Nico, Matthieu Caruel, et bien sur Gérard, pour ces funky midis et ces barbecues enflammés ! Merci aussi aux LMS-girls : Waad, Camille et Shadan et au capitaine de l’OLMS, Fabien, pour leur bonne humeur permanente. Enfin, merci aux habitants du bâtiment 65 pour leur accueil chaleureux : Emmanuel, Pascal, Abdul, Jean-Marc, Franz, Yongyun, Christelle, Diego & Benoit. De façon globale merci aux membres du pôle *Mécanique pour les vivants* ainsi qu’à ceux du LMS pour l’accueil qu’il m’ont offert.

En dehors du laboratoire, c’est surtout ma famille qui m’a permis de garder ma motivation intacte. Même si mon travail, et surtout ce que je “*cherchais*”, vous a semblé obscur (voire fumeux ...) jusqu’au jour de la soutenance, vous vous êtes finalement rendu compte que je ne passais pas uniquement mon temps à jouer avec des *wiimotes* en écoutant de la musique. Merci donc pour votre soutien durant ma thèse et promis, maintenant je vais tâcher de trouver un “*vrai*” travail. Surtout, je tiens particulièrement à remercier Caroline, ma femme, pour ses acquiescements lorsque j’avais des *idées* par rapport au travail à la maison ou dans la voiture, pour sa relecture et ses commentaires concernant le manuscrit de thèse, et pour son soutien sans faille durant ces trois ans. Pour conclure, merci à Gabriel qui, de par son arrivée, m’a imposé sans le vouloir une *deadline* non négociable à laquelle la thèse devait être finie.

Résumé

La réalité virtuelle ouvre une fenêtre, voulue transparente, sur un monde virtuel dans lequel sont plongés des participants. Conçus d’abord comme uniquement *visuels*, les mondes virtuels sont maintenant aussi *sonores*, grâce aux technologies récentes de rendu sonore spatialisé. Le système innovant proposé et étudié ici se compose de larges panneaux de structure sandwich, servant d’écrans, au revers desquels sont collés des excitateurs électrodynamiques qui les mettent en vibration, créant ainsi un champ acoustique autour des auditeurs. Le rendu visuel 3D est généré par stéréoscopie passive adaptative et le rendu sonore spatialisé par holophonie.

L’utilisation d’un panneau comme source sonore étendue pour l’holophonie suppose la compréhension de son comportement vibro-acoustique. L’estimation des non-linéarités dans les systèmes vibrants est abordée à partir d’une méthode proposée par A. Farina en 2000 : cette méthode est justifiée sur le plan théorique et son applicabilité est étendue (séparation des sources de non-linéarité). Une méthode d’estimation des *modules d’élasticité* et des *facteurs d’amortissement* pertinents pour la propagation dans le panneau sandwich des ondes de flexion et de cisaillement est aussi développée. Le système est caractérisé expérimentalement grâce à ces deux outils.

La perception spatiale, par les modalités audio, visuelle et audio-visuelle, du monde virtuel synthétisé par le dispositif est ensuite abordée. Il est montré que les participants fusionnent avec succès les flux auditif et visuel qui leur sont présentés et perçoivent les rendus auditif et visuel de façon spatialement cohérente. Les caractéristiques de leur perception spatiale de l’espace virtuel sont analysées en fonction des modalités mises en jeu et comparées avec les données de la littérature. Le concept d’*ancrage* du monde virtuel au dispositif réel qui le crée est en outre proposé : la distance entre le sujet et le panneau semble en effet jouer un rôle pivot dans la loi de compression des distances perçues.

Enfin, deux applications de réalité virtuelle basées sur ce système illustrent les nombreuses possibilités offertes par cette nouvelle interface “homme-machine”.

Mots-clés : Réalité virtuelle, son spatialisé, non-linéarités, matériau sandwich, perception spatiale, intégration audio-visuelle.

Abstract

Virtual reality provides a transparent window open into a virtual world in which users are immersed. Initially being primarily *visual*, these virtual worlds now also include *sound*, thanks to recently developed spatialized sound rendering technologies. A new system proposed and studied here is comprised of large sandwich panels, used as screens, on the back of which electro-dynamic exciters are attached, thus allowing for the creation of an acoustic field surrounding the users. The 3D visual rendering is generated using tracked passive stereoscopy and the spatial sound rendering is achieved by means of holophony, also known as wave-field synthesis.

Using a sandwich panel for holophony supposes that its vibro-acoustic behavior is well known. The estimation of non-linearities in vibrating systems is addressed on the basis of an empirical method proposed by A. Farina in 2000. The theoretical basis of the method is justified and extended to allow also for the separation of the contributions of different non-linear components in the measurement chain. A method for the estimation of *elastic* and *damping* factors, those that are relevant to the propagation of flexural and shear waves in the sandwich panel, has also been developed. Thanks to these two tools, the system under study has been experimentally characterized.

Spatial perception of the virtual world synthesized by the system, using solely audio, visual, or both modalities is then studied. It is shown that participants successfully merge the presented audio and visual streams into a single fused object. Specifically, the audio and visual rendering are perceived as spatially coherent. The characteristics of spatial perception of the virtual world are analyzed with respect to the rendering modalities and results are compared with previous studies. The concept of *anchorage* of the virtual world to the real world is furthermore proposed: the physical distance between the user and the system affects the perceived degree of distance compression for both modalities.

Finally, two virtual reality applications illustrate some of the numerous possibilities offered by this new “computer/human interface”.

Keywords: Virtual reality, spatialized sound rendering, non-linearity, sandwich materials, spatial perception.

Table des matières

1	Introduction générale	19
1.1	De la vibration à la perception d'un environnement virtuel	19
1.2	Organisation du manuscrit	20
1.2.1	Son spatialisé et réalité virtuelle	21
1.2.2	Comportement dynamique des LaMAPs comme sources sonores pour l'holophonie	21
1.2.3	Perception spatiale du monde virtuel synthétisé par le SMART-I ²	22
2	Intégration d'un rendu sonore spatialisé dans un système audio-visuel de réalité virtuelle	23
2.1	Introduction	24
2.2	Réalité virtuelle et transparence d'une interface	24
2.2.1	Réalité virtuelle	24
2.2.2	Présence	25
2.2.3	Transparence d'une interface et immersion	26
2.2.4	Importance du contenu	27
2.3	Rendu sonore spatialisé et intérêt en environnement audio-visuel virtuel	28
2.3.1	Rendu sonore spatialisé	28
2.3.2	Impact sur la sensation de présence	29
2.3.3	Impact sur les réactions émotionnelles	31
2.3.4	Impact sur la compréhension d'environnements complexes	32
2.3.5	Impact sur la <i>qualité globale audio-visuelle</i> perçue	33
2.3.6	Bilan	34
2.4	Intégration d'un système de rendu sonore spatialisé au sein d'une interface visuelle de réalité virtuelle typique	35
2.4.1	Une interface visuelle de réalité virtuelle typique : la <i>CAVE</i>	36
2.4.2	Technologies de création de rendu sonore spatialisé	37
2.4.3	Choix d'une technologie pour l'intégration d'un <i>rendu sonore spatialisé</i> avec une interface visuelle de type <i>CAVE</i>	38
2.5	Le SMART-I ²	42
2.5.1	Spécificités du système	42
2.5.2	Vue d'ensemble	43
2.5.3	Rendu sonore spatialisé par « <i>Wave Field Synthesis</i> »	44
2.5.4	Les <i>Large Multi-Actuator Panels</i> (LaMAPs)	49

2.5.5	Stéréoscopie passive avec suivi	51
2.6	Conclusion	55
3	Applications de réalités virtuelles basées sur le SMART-I²	57
3.1	Introduction	57
3.2	GAVIP : “ <i>Gestural Auditory and Visual Interactive Platform</i> ”	58
3.2.1	Concept et motivations du projet	58
3.2.2	Architecture logicielle mise en place	59
3.2.3	Un premier scénario d’interaction : « <i>Bubulles</i> »	60
3.2.4	Vers un second scénario : « <i>Lava Lova</i> »	61
3.2.5	Perspectives	63
3.3	Intégration de <i>MARC</i> (Multi-modal Affective Reactive Characters) dans le SMART-I ² . . .	64
3.3.1	La plateforme <i>MARC</i> (Multi-modal Affective Reactive Characters)	64
3.3.2	Scène de démonstration développée pour les 5ièmes journées de l’Association Française de Réalité Virtuelle (AFRV)	65
3.3.3	Premier retour « <i>utilisateur</i> » à propos du rendu audio-visuel proposé par le SMART-I ² . . .	66
3.3.4	Perspectives	68
3.4	Conclusion	68
4	Mesure des non-linéarités dans les systèmes vibrants	69
4.1	Introduction	69
4.2	Identification de modèles de Hammerstein en cascade pour la description des non-linéarités dans les systèmes vibrants	70
4.2.1	Introduction	71
4.2.2	Modélisation et mesure des non-linéarités	72
4.2.3	Bases mathématiques de la méthode	75
4.2.4	Implémentation pratique	78
4.2.5	Validation de la méthode	80
4.2.6	Modélisation de transducteurs acoustiques par des modèles de Hammerstein en cascade	84
4.2.7	Prédiction de la distorsion harmonique générée par les transducteurs	87
4.2.8	Discussion	91
4.2.9	Conclusion	91
4.3	Discussion	93
4.3.1	Effets de différents paramètres sur la qualité de l’estimation des noyaux	93
4.3.2	Comparaison avec les travaux d’A. Novák <i>et al.</i>	95
4.3.3	Chaîne de systèmes non-linéaires	101
4.3.4	Librairie Matlab <i>Hammerstein Toolbox</i>	107
4.4	Conclusion et perspectives	107

5	Mesure dynamique des propriétés d'élasticité et de dissipation des constituants de plaques "sandwich"	109
5.1	Introduction	110
5.2	Étude qualitative du comportement dynamique d'une tranche de matériau sandwich symétrique	110
5.2.1	Définition d'un matériau de type "sandwich" symétrique	110
5.2.2	Types d'ondes existant dans un matériau sandwich symétrique	110
5.2.3	Modèles dynamiques élémentaires	112
5.2.4	Courbes de dispersion	114
5.2.5	Comportement asymptotique en basses fréquences	116
5.3	Détermination de paramètres d'élasticité et d'amortissement pertinents pour la modélisation de plaques "sandwich" épaisses	117
5.3.1	Introduction	117
5.3.2	Un modèle mécanique des panneaux "sandwich"	119
5.3.3	Estimation des paramètres modaux par une procédure de Rayleigh-Ritz étendue	124
5.3.4	L'analyse modale haute-résolution pour l'estimation des paramètres modaux	126
5.3.5	Procédure d'optimisation	129
5.3.6	Validation de la procédure d'estimation	131
5.3.7	Résultats expérimentaux	135
5.3.8	Conclusion	142
5.4	Discussion des hypothèses sur le rayonnement acoustique adoptées pour l'étude dynamique des panneaux	144
5.4.1	Hypothèse de <i>fluide léger</i>	144
5.4.2	Hypothèse de <i>faible amortissement</i>	148
5.4.3	Majoration des pertes par rayonnement	152
5.5	Conclusion	155
6	Caractérisation expérimentale des "Large Multi-Actuator Panels" (LaMAPs)	157
6.1	Introduction	157
6.2	Modélisation simple des LaMAPS	158
6.2.1	Modèle de plaque mince homogène, isotrope et infinie	158
6.2.2	Grandeurs physiques caractéristiques de la dynamique des LaMAPs	158
6.2.3	Action des excitateurs sur le panneau	160
6.3	Confrontation du modèle à l'expérience	161
6.3.1	Protocole expérimental	161
6.3.2	Analyse temporelle	163
6.3.3	Analyse fréquentielle	169
6.4	Conclusion et perspectives	173

7	Cohérence audio-visuelle de l'espace virtuel synthétisé par le SMART-I²	175
7.1	Introduction	175
7.2	Perception spatiale unimodale	176
7.2.1	Espace perceptif auditif	176
7.2.2	Espace perceptif visuel	182
7.3	Intégration spatio-temporelle audio-visuelle	184
7.3.1	Fenêtre spatio-temporelle d'intégration audio-visuelle	185
7.3.2	Facteurs cognitifs influents	188
7.3.3	Conséquences	189
7.4	Évaluation des performances offertes par le SMART-I ²	190
7.4.1	Présentation générale	190
7.4.2	Analyse objective	192
7.4.3	Précision de localisation en azimuth	194
7.4.4	Restitution de l'effet de parallaxe	198
7.4.5	Effet de l'apprentissage	202
7.5	Conclusion	203
8	Transparence de l'interface de restitution et perception de l'espace virtuel	205
8.1	Introduction	205
8.2	Distance égocentrique audio, visuelle et audio-visuelle perçue en environnement virtuel par des sujets mobiles	206
8.2.1	Introduction	206
8.2.2	Méthode	209
8.2.3	Analyse des résultats	213
8.2.4	Discussion	219
8.2.5	Conclusion	225
8.3	Discussion	226
8.3.1	Retour sur l'intégration audio-visuelle en distance	226
8.3.2	Limites de la procédure de triangulation (étude préliminaire)	228
8.3.3	Comparaison des deux expériences d'estimation de la distance audio	230
8.4	Conclusion	232
9	Conclusion générale	235
9.1	Contributions de la thèse	235
9.1.1	Son spatialisé et réalité virtuelle	236
9.1.2	Comportement dynamique des LaMAPs comme sources sonores pour l'holophonie	236
9.1.3	Perception spatiale du monde virtuel synthétisé par le SMART-I ²	237
9.2	Quelques perspectives	238
9.2.1	Améliorations possibles du SMART-I ²	238
9.2.2	Impact d'un rendu sonore spatialisé sur des applications de réalité virtuelle	238
9.2.3	Non-linéarités générées par un système vibrant	239

TABLE DES MATIÈRES

9.2.4	Effets des inhomogénéités introduites par les nids d'abeille	239
9.2.5	Modélisation des LaMAPs en vue d'un contrôle vibratoire pour l'holophonie . . .	240
9.2.6	Variations autour du protocole expérimental de la section 8.2	240
9.3	Publications liées à la thèse	241
9.3.1	Articles	241
9.3.2	Acte de conférences	241
A	Annexe “Non-linéarités” : Calcul de la matrice C	243
B	Détails concernant la procédure d'estimation des propriétés d'élasticité et de dissipation des constituants de plaques “sandwich”	245
B.1	Génération de la base polynomiale orthonormale	245
B.2	Expressions analytiques des dérivées de T et U	245
B.3	Expressions analytiques de t_n et u_k^n	246
B.4	Nomenclature de la section 5.3	247
	Bibliographie	251

Chapitre 1

Introduction générale

Sommaire

1.1	De la vibration à la perception d'un environnement virtuel	19
1.2	Organisation du manuscrit	20
1.2.1	Son spatialisé et réalité virtuelle	21
1.2.2	Comportement dynamique des LaMAPs comme sources sonores pour l'holophonie	21
1.2.3	Perception spatiale du monde virtuel synthétisé par le SMART-I ²	22

1.1 De la vibration à la perception d'un environnement virtuel

L'objectif d'une application de réalité virtuelle audio-visuelle spatialisée est d'immerger un participant dans un monde virtuel audio-visuel en lui proposant une fenêtre ouverte sur cet espace synthétique. Pour que le participant se sente *présent* dans le monde virtuel, il est impératif de garantir que cette fenêtre audio-visuelle soit complètement *transparente* d'un point de vue perceptif.

Cependant, les systèmes de réalité virtuelle les plus courants, *i.e.* les systèmes de type *CAVE*¹, souffrent de défauts qui nuisent à la *transparence*. La présence d'écrans de projection crée en effet un environnement acoustique peu adapté à la restitution sonore spatialisée. De plus, les conflits de positionnement entre écrans, vidéo-projecteurs et haut-parleurs empêchent la distribution optimale des haut-parleurs dans l'espace de restitution. Ces deux facteurs dégradent donc la qualité spatiale du rendu sonore final.

Créer un rendu sonore spatialisé de qualité dans un système audio-visuel est pourtant un enjeu de taille pour les applications de réalité virtuelle. En effet, les modalités audio et visuelle n'étant pas découplées l'une de l'autre, leur combinaison cohérente dans l'espace est potentiellement source d'améliorations. Entre autres, l'ajout d'un rendu sonore spatialisé à un système de réalité virtuelle impacte favorablement la sensation de *présence* et la *qualité* audio-visuelle ressenties par les participants, ainsi que leur capacité de *compréhension* de scènes audio-visuelles complexes (voir par exemple [Bormann 2005, Kohlrausch & van de Par 2005, Gröhn et al. 2005]).

Fort de ce constat, le SMART-I² ("*Spatial, Multi-user, Audio-visuel, Real-Time, Interactive Interface*") a été développé durant cette thèse dans le but d'intégrer de façon optimale un système de rendu sonore

1. *CAVE* : Audio Visual Experience Automatic Virtual Environment, d'après Cruz-Neira et al. [1992]

spatialisé à un système de réalité virtuelle de type *CAVE*. La spécificité de ce dispositif est de fusionner les écrans de projection et les haut-parleurs au sein d'un seul et unique élément technologique, nommé "*Large Multi-Actuator Panel*" (LaMAP). Un LaMAP est constitué d'un panneau de matériau sandwich de grandes dimensions au dos duquel sont collés des excitateurs électro-dynamiques. Sa face avant sert d'écran de projection visuel. Les excitateurs électro-dynamiques situés sur sa face arrière engendrent des ondes de flexion qui rayonnent acoustiquement dans l'espace de restitution. Le rendu visuel tridimensionnel est généré par *stéréoscopie passive adaptative* et le rendu sonore spatialisé par *holophonie* ("*Wave Field Synthesis*" [Berkhout et al. 1993]). Le SMART-I² permet donc de s'affranchir des défauts liés à la présence d'écrans de projection ainsi que des conflits de positionnement empêchant la distribution optimale des haut-parleurs dans l'espace de restitution. Ce dispositif garantit donc une grande cohérence spatiale audio-visuelle tout en ne faisant que très peu de compromis sur les qualités spatiales audio et visuelle : la fenêtre audio-visuelle proposée est a priori "*très*" transparente.

Ainsi, les vibrations de panneaux de grandes dimensions faits de matériau sandwich sont à la base du rendu sonore spatialisé proposé par le SMART-I². À l'aide de ces vibrations, un espace sonore cohérent avec un espace visuel synthétique est créé par holophonie. Le tout est destiné à des applications de réalité virtuelle. La problématique de cette thèse est donc : comment transformer des *vibrations* en indices sonores spatiaux *perceptibles* d'un *environnement virtuel* ?

L'objectif de cette thèse est alors d'apporter des éléments de réponse à cette problématique en l'abordant selon trois axes principaux :

- **Son spatialisé et réalité virtuelle**
 - Pourquoi et comment intégrer un rendu sonore spatialisé dans un système de type "*CAVE*" ?
 - Comment évaluer l'impact de l'ajout d'un système de rendu sonore spatialisé ?
- **Comportement dynamique des LaMAPs utilisés comme sources sonores pour l'holophonie**
 - Comment caractériser les non-linéarités générées par un système vibro-acoustique ?
 - Comment modéliser le comportement dynamique d'un matériau sandwich ?
 - Quelle est, en pratique, le comportement dynamique d'un LaMAP ?
- **Perception spatiale du monde virtuel synthétisé par le SMART-I²**
 - Le monde virtuel audio-visuel créé par le SMART-I² est-il perceptivement spatialement cohérent ?
 - Quel est l'impact de la transparence limitée du SMART-I² sur l'espace virtuel perçu ?

1.2 Organisation du manuscrit

Le manuscrit présenté ici est structuré autour des différentes questions rattachées aux trois axes listés précédemment. Les interrogations relatives au son spatialisé et à la réalité virtuelle sont abordées dans les chapitres 2 et 3. Le comportement vibro-acoustique des LaMAPs, utilisés en tant que sources sonores étendues pour l'holophonie, est ensuite étudié dans les chapitres 4, 5 et 6. Enfin, la perception de l'espace audio-visuel virtuel synthétisé par le SMART-I² fait l'objet des chapitres 7 et 8. L'organisation thématique du manuscrit est schématiquement représentée sur la figure 1.1. Le contenu des trois axes principaux est détaillé dans les sections 1.2.1, 1.2.2 et 1.2.3.

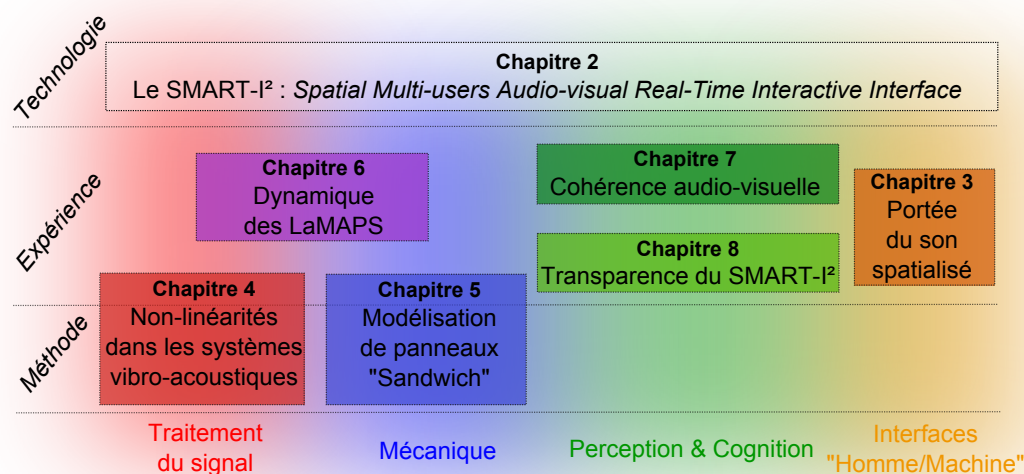


FIG. 1.1 – Organisation thématique transversale des différents chapitres composant la thèse.

1.2.1 Son spatialisé et réalité virtuelle

Le chapitre 2 présente le SMART-I² ainsi que son contexte applicatif : la réalité virtuelle. Dans un premier temps, la notion de *réalité virtuelle* est introduite en mettant l'accent sur la spécificité de l'approche à adopter lors de la conception d'un système destiné à ce contexte. Puis le concept de *rendu sonore spatialisé*, caractérisé par quatre attributs : la *direction*, la *distance*, la *réverbération* et l'*effet de parallaxe*, est proposé et une étude bibliographique analyse l'impact que peut avoir un *rendu sonore spatialisé* sur différents aspects de la perception d'un environnement audio-visuel de réalité virtuelle. Les compromis technologiques à réaliser lors de l'ajout d'un dispositif de *rendu sonore spatialisé* à une interface visuelle pour la réalité virtuelle de type "CAVE" sont aussi analysés.

Le chapitre 3 décrit ensuite deux projets en cours dont l'objectif à long terme est l'étude de l'impact des différents attributs d'un *rendu sonore spatialisé* sur la sensation de *présence* perçue. L'un des projets propose une immersion *active* des participants dans le monde virtuel (GAVIP), tandis que l'autre propose une immersion plutôt *passive* (MARC & SMART-I²).

1.2.2 Comportement dynamique des LaMAPs comme sources sonores pour l'holophonie

Le chapitre 4 traite de l'estimation des non-linéarités générées par les systèmes faiblement non-linéaires. La première partie est constituée d'un article publié dans le *Journal of Sound and Vibration* intitulé "*Identification of cascade of Hammerstein models for the description of non-linearities in vibrating devices*" qui présente une méthode originale, inspirée des travaux empiriques d'Angelo Farina [Farina 2000, Farina et al. 2001]. Cette méthode permet d'estimer à l'aide d'une seule mesure un modèle de système non-linéaire correspondant à un système vibrant. La seconde partie de ce chapitre présente une discussion autour de cette

nouvelle méthode : l'influence de différents paramètres sur la qualité des estimations est discutée, puis la méthode développée est comparée aux travaux d'A. Novák et al. [2010a], et le cas d'une chaîne de systèmes non-linéaires est considéré. Enfin, une boîte à outils Matlab disponible en ligne et proposant une implémentation de la méthode proposée est brièvement introduite.

Le chapitre 5 présente et valide une méthode permettant l'estimation des paramètres mécaniques nécessaires à la description de la dynamique des panneaux de matériau sandwich composant le SMART-I². Une étude qualitative de la dynamique des matériaux sandwich décrivant les différents types d'ondes susceptibles de se propager dans un tel matériau est d'abord proposée. Puis, sous la forme d'un article accepté au *Journal of Sound and Vibration*, sont présentés la méthode proprement dite, une validation basée sur un modèle "*éléments finis*", et les résultats expérimentaux obtenus sur différents échantillons. Enfin, le domaine de validité des différentes hypothèses émises concernant le rayonnement acoustique est discuté.

Le chapitre 6 présente enfin une étude préliminaire apportant un éclairage expérimental sur différents phénomènes influençant la dynamique des LaMAPs (les conditions aux limites, la présence des excitateurs et les interactions entre les deux panneaux) en comparant des résultats expérimentaux obtenus sur les LaMAPs aux prédictions d'un modèle dynamique simple (modèle de plaque mince, homogène, isotrope, infinie avec excitation ponctuelle).

1.2.3 Perception spatiale du monde virtuel synthétisé par le SMART-I²

L'objectif du chapitre 7 est de déterminer dans quelle mesure le SMART-I² permet à un utilisateur, plongé dans l'environnement de réalité virtuelle qu'il synthétise, d'utiliser ses capacités perceptives audio-visuelles et ses capacités motrices pour comprendre l'espace virtuel et se déplacer en son sein. Dans ce but, la façon dont l'espace est perçu au travers des modalités auditive et visuelle est d'abord rappelée, puis les conditions perceptives requises pour que les flux spatio-temporels d'informations provenant des modalités auditive et visuelle fusionnent et donnent naissance à un unique flux audio-visuel sont décrites. Deux expériences, ayant pour but d'estimer la précision de localisation en azimuth, d'étudier la capacité de restitution de l'effet de parallaxe auditive, ainsi que d'évaluer la cohérence perceptive du rendu audio-visuel proposé par le SMART-I² sont ensuite présentées.

Le chapitre 8 s'attache enfin à comprendre comment la perception de l'environnement audio-visuel synthétisé par le SMART-I² est liée à la connaissance que l'on a du dispositif physique. Plus précisément, une expérience centrée sur l'influence de la *transparence* limitée du SMART-I² sur la perception de la distance et sur la sensation de *présence* est présentée. Cette partie a fait l'objet d'une soumission au journal *Transactions on Applied Perception*.

Chapitre 2

Intégration d'un rendu sonore spatialisé dans un système audio-visuel de réalité virtuelle

Sommaire

2.1	Introduction	24
2.2	Réalité virtuelle et transparence d'une interface	24
2.2.1	Réalité virtuelle	24
2.2.2	Présence	25
2.2.3	Transparence d'une interface et immersion	26
2.2.4	Importance du contenu	27
2.3	Rendu sonore spatialisé et intérêt en environnement audio-visuel virtuel	28
2.3.1	Rendu sonore spatialisé	28
2.3.2	Impact sur la sensation de présence	29
2.3.3	Impact sur les réactions émotionnelles	31
2.3.4	Impact sur la compréhension d'environnements complexes	32
2.3.5	Impact sur la <i>qualité globale audio-visuelle</i> perçue	33
2.3.6	Bilan	34
2.4	Intégration d'un système de rendu sonore spatialisé au sein d'une interface visuelle de réalité virtuelle typique	35
2.4.1	Une interface visuelle de réalité virtuelle typique : la <i>CAVE</i>	36
2.4.2	Technologies de création de rendu sonore spatialisé	37
2.4.3	Choix d'une technologie pour l'intégration d'un <i>rendu sonore spatialisé</i> avec une interface visuelle de type <i>CAVE</i>	38
2.5	Le SMART-I²	42
2.5.1	Spécificités du système	42
2.5.2	Vue d'ensemble	43
2.5.3	Rendu sonore spatialisé par « <i>Wave Field Synthesis</i> »	44
2.5.4	Les <i>Large Multi-Actuator Panels</i> (LaMAPs)	49

2.5.5 Stéréoscopie passive avec suivi	51
2.6 Conclusion	55

2.1 Introduction

Le premier chapitre de cette thèse présente le SMART-I² («*Spatial Multi-user Audio-visuel Real-Time Interactive Interface*»), un dispositif de restitution audio-visuelle spatialisée qui a été développé pour des applications de réalité virtuelle au cours de cette thèse.

Dans un premier temps, la *réalité virtuelle* est définie (section 2.2.1). Puis la notion de *rendu sonore spatialisé* est établie et une étude bibliographique analyse l'impact que peut avoir un *rendu sonore spatialisé* sur différents aspects de la perception d'un environnement audio-visuel de réalité virtuelle (section 2.3). Les compromis technologiques à réaliser lors de l'ajout d'un dispositif de *rendu sonore spatialisé* à une interface visuelle pour la réalité virtuelle sont ensuite analysés (section 2.4). Motivé par l'intérêt d'ajouter un *rendu sonore spatialisé* pour des applications de *réalité virtuelle* tout en cherchant à éviter les compromis technologiques, le système SMART-I² est enfin décrit (section 2.5).

2.2 Réalité virtuelle et transparence d'une interface

Dans cette section, la notion de *réalité virtuelle* est introduite en mettant l'accent sur la spécificité de l'approche à adopter lors de la conception d'un système destiné à la réalité virtuelle, en particulier en regard des notions de *transparence*, d'*immersion* et de *présence*. L'importance du contenu proposé par les applications de réalité virtuelle est aussi soulignée.

2.2.1 Réalité virtuelle

Dans les représentations *figées* de mondes réels ou imaginaires (peintures, photographies, enregistrements audio, etc ...) l'utilisateur est simplement *spectateur* et non *acteur* au sein de l'environnement qui lui est proposé. La *réalité virtuelle* peut alors être vue comme la science permettant à l'homme de passer du statut de *spectateur* à celui d'*acteur* au sein de ces représentations, qu'elles soient réalistes ou non. La définition suivante de la *réalité virtuelle* découle de cette idée [Fuchs et al. 2006b] :

«La *réalité virtuelle* est le domaine scientifique et technique exploitant l'informatique et des interfaces comportementales en vue de simuler dans un monde virtuel le comportement d'entités **autonomes**¹, qui sont en **interaction** en temps réel entre elles et avec un ou des utilisateurs en **immersion** pseudo-naturelle par l'intermédiaire de canaux sensori-moteurs.»

Par définition, la *réalité virtuelle* se positionne donc comme un domaine scientifique *multi-disciplinaire*, mettant en jeu sciences humaines et sciences de l'ingénieur. De plus, il émerge de la définition précédente trois critères permettant de situer la *réalité virtuelle* par rapport à d'autres applications : l'*autonomie* du monde virtuel par rapport à l'utilisateur, l'*interaction* entre l'utilisateur et le monde virtuel, et l'*immersion* de l'utilisateur dans le monde virtuel.

1. Le terme «*autonome*» n'apparaît pas dans la définition originale donnée par Fuchs et al. [2006b]. Il a été introduit ici pour concilier l'approche présentée dans Fuchs et al. [2006b] avec celle de Tisseau [2004].

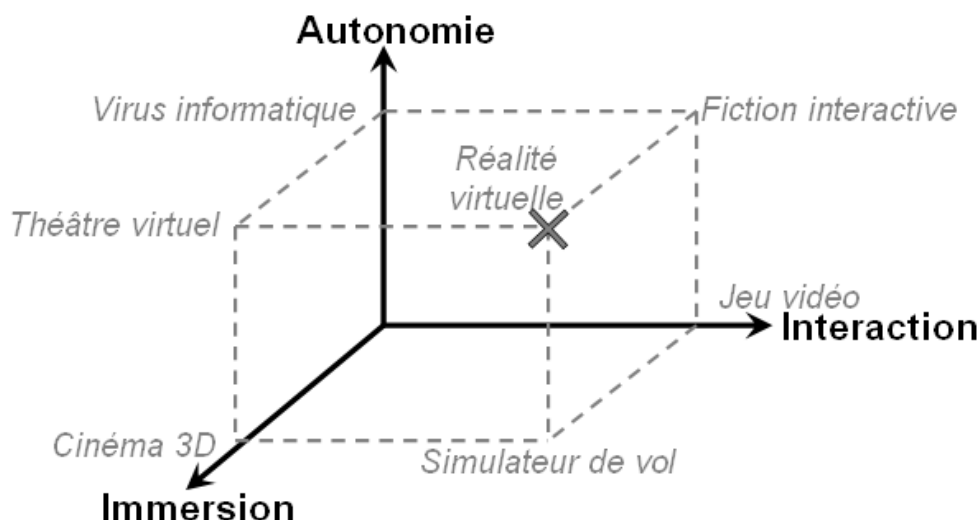


FIG. 2.1 – La réalité virtuelle située parmi d'autres applications dans l'espace «autonomie/interaction/immersion» (d'après Tisseau [2004]).

Sur la figure 2.1, plusieurs applications sont représentées par un point dans l'espace *immersion / interaction / autonomie* [Tisseau 2004]. Dans ce repère, le cinéma 3D correspond à une application où l'accent est mis sur l'*immersion*, le jeu-vidéo à une application où domine l'*interaction*, alors qu'un simulateur de vol classique propose à l'utilisateur à la fois *immersion* et *interaction*. Le virus informatique symbolise l'*autonomie* car il échappe à celui qui l'a créé sans être maîtrisé par l'utilisateur. Le théâtre virtuel permet à un utilisateur d'être *immergé* en tant qu'observateur, libre de se déplacer dans une scène jouée par des acteurs virtuels *autonomes*, mais sans pouvoir *interagir* avec eux. À l'inverse, la fiction interactive (jeu de rôle) permet à l'utilisateur *non immergé* d'*interagir* avec des acteurs *autonomes*.

Une application de réalité virtuelle peut être donc généralement définie comme *une application autorisant un utilisateur immergé à interagir avec des entités virtuelles autonomes*.

2.2.2 Présence

La finalité des applications de réalité virtuelle, telles que formalisées dans la section 2.2.1, est de plonger l'utilisateur dans un monde virtuel où il agit et réagit *comme s'il était dans le monde réel*. Il est donc impératif de créer des systèmes de réalité virtuelle permettant aux utilisateurs de se sentir *présents* dans l'environnement virtuel. Ce sentiment est décrit par le concept de *présence* [Lombard & Ditton 1997]. La définition de la *présence* comme «*sentiment d'être dans le monde virtuel*» semble être consensuelle, au moins pour des applications de réalité virtuelle où l'accent est mis sur l'*immersion* [Lombard & Ditton 1997, Schuemie et al. 2001, Sanchez-Vives & Slater 2005]. On remarque de plus que la notion de *présence* est assez similaire à la notion de *validité écologique*, spécifiant qu'un protocole expérimental (qui n'est rien d'autre qu'un environnement virtuel particulier) est *écologiquement valide* si «*les participants réagissent [à lui], dans une certaine mesure, comme s'ils étaient dans une situation naturelle*» [Guastavino et al. 2005].

D'un point de vue théorique, la *présence* regroupe les dimensions «*immersion*» et «*interaction*» décrites

dans la section 2.2.1 [Tisseau 2004]. Il est par ailleurs supposé qu'un utilisateur se sentira complètement *présent* dans un environnement virtuel seulement en cas d'*équivalence cognitive et sensorielle* entre le monde virtuel et le monde réel [Lombard & Ditton 1997]. L'*équivalence sensorielle* est possible à la condition que les dispositifs de rendus reproduisent fidèlement dans l'environnement virtuel les flux sensoriels qui seraient perçus dans le monde réel. L'*équivalence cognitive* suppose que le *modèle mental* que se fait l'utilisateur du monde virtuel est équivalent à celui qu'il se ferait du monde réel en interagissant avec un environnement comportant les mêmes objets.

Les facteurs influents sur la *présence* ressentie par les utilisateurs, d'un point de vue sensoriel ou cognitif, ainsi que les effets de la *présence* sur les actions des utilisateurs ou sur leur représentation mentale du monde virtuel sont nombreux et restent encore mal connus [Lombard & Ditton 1997, Schuemie et al. 2001, Sanchez-Vives & Slater 2005]. Ceci est principalement dû au fait qu'il est difficile de mesurer objectivement la *présence* [IJsselsteijn et al. 2000, Schuemie et al. 2001, Insko 2003]. Les outils les plus courants utilisés pour obtenir une mesure de la *présence* ressentie sont des questionnaires post-expérimentaux. Les plus utilisés sont le questionnaire *Slater-Usch-Steer* (SUS) de Slater & Usch [1994], le *Presence Questionnaire* (PQ) élaboré par Witmer & Singer [1998], l'*ITC-Sense of Presence Inventory* (ITC-SOPI) de Lessiter & Freeman [2001] et l'*Igroup Presence Questionnaire* (IPQ) [Schubert et al. 2001].

La *présence* est donc un facteur clé en réalité virtuelle et constitue une des quantités à optimiser lors de la création d'un dispositif destiné à la réalité virtuelle. Cependant, les facteurs facilitant la sensation de *présence* restent mal connus.

2.2.3 Transparence d'une interface et immersion

Selon l'approche théorique de la *présence* décrite à la section 2.2.2, l'*équivalence sensorielle* implique que les dispositifs de restitution du monde virtuel soient *transparents* par rapport à l'utilisateur (voir figure 2.2). L'utilisateur doit en effet percevoir le monde virtuel qui lui est proposé et non l'interface par laquelle il perçoit le monde virtuel. Une métaphore permettant d'illustrer ceci est présentée par Slater & Wilbur [1997] : «*Pour les personnes assises dans une voiture, la réalité extérieure apparaît distante et étrangère, placée de l'autre côté d'une barrière ou d'un écran matérialisé par la vitre. L'extérieur est alors perçu comme une autre réalité, une réalité qui n'est pas directement contiguë à la réalité de l'intérieur de la voiture. La preuve de cette discontinuité est l'inconfort généralement ressenti lors de l'ouverture de la fenêtre et de l'irruption de la réalité extérieure dans celle de l'intérieur de la voiture.*» Une interface de réalité virtuelle doit donc être vue comme une fenêtre à travers laquelle un monde virtuel est perçu, cette fenêtre devant être la plus *transparente* possible pour que le monde virtuel devienne un prolongement naturel du monde réel [Brooks 1999].

Or il est évident que quelles que soient les technologies de restitution utilisées, leur fidélité par rapport à la réalité sera toujours limitée et leur *transparence* le sera par conséquent aussi (latence dans le rendu, champ de vision limité par un écran, port d'un casque audio intrusif, etc ...). Il est cependant possible de chercher à réduire l'écart sensoriel entre les mondes réels et virtuels en utilisant des technologies sophistiquées et en stimulant plusieurs modalités sensorielles à la fois. Toujours en cohérence avec l'approche générale de la section 2.2.1, on peut alors définir l'*immersion* comme «*l'ensemble des moyens technologiques effectivement mis en œuvre pour plonger l'utilisateur dans le monde virtuel*» [Slater & Wilbur

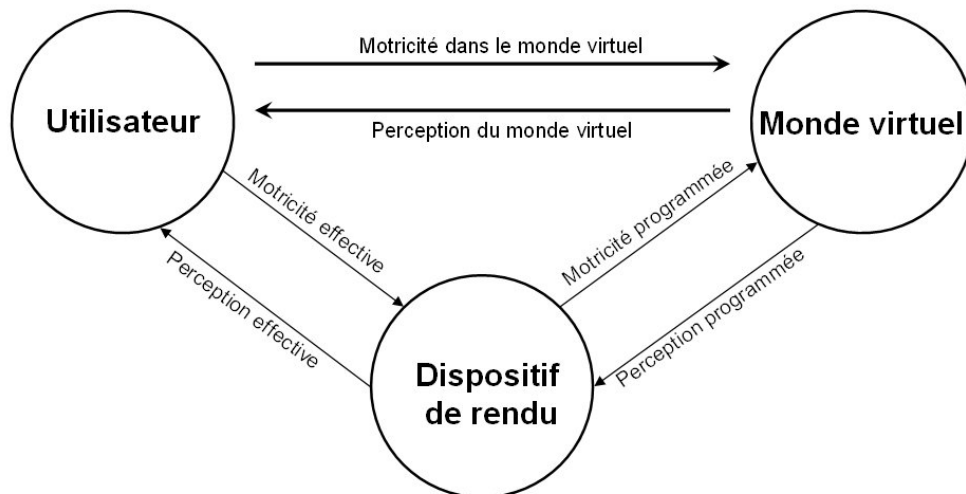


FIG. 2.2 – Illustration de la notion de *transparence* d'un dispositif de rendu en réalité virtuelle (d'après [Fuchs et al. 2006b]).

1997]. L'*immersion* est donc la combinaison des critères objectifs permettant de caractériser un dispositif de réalité virtuelle.

À la différence de la *présence* qui est par essence subjective, l'*immersion* est quantifiable et est donc un facteur objectif [Slater & Wilbur 1997, Slater 2003, Sanchez-Vives & Slater 2005, Bowman & McMahan 2007]. Même si la relation entre *immersion* et *présence* n'est pas clairement établie, il a été montré que l'*immersion* est un des facteurs influençant la *présence* [Lombard & Ditton 1997, IJsselsteijn et al. 2000, Lessiter & Freeman 2001, Schuemie et al. 2001]. Il n'est cependant pas garanti qu'une augmentation de l'*immersion* conduise à une augmentation de la *présence* [Sanchez-Vives & Slater 2005, Bowman & McMahan 2007].

Ainsi, une interface destinée à des applications de réalité virtuelle devra principalement chercher à maximiser la *présence* subjectivement perçue par les sujets. Une façon d'atteindre cet objectif est de proposer une interface qui soit la plus *transparente* possible. Il est cependant primordial de s'assurer que les efforts faits pour augmenter l'*immersion* aient un impact bénéfique sur la *présence* subjectivement perçue².

2.2.4 Importance du contenu

Au même titre que l'*immersion*, décrivant la façon dont un environnement virtuel est exposé aux participants, le *contenu de l'environnement virtuel* (objets, êtres vivants, etc.), et la possibilité qu'aura le participant d'interagir avec lui (*tâche effectuée*) impacteront directement la sensation de *présence*. Les deux facteurs les plus influents sur la *présence*, du point de vue du *contenu*, sont ainsi [Lombard & Ditton 1997, Lessiter & Freeman 2001] :

2. Par exemple, Guastavino & Katz [2004] montrent que, dans un contexte audio uniquement, les systèmes de restitution sonore 2D (*i.e.* n'ayant des haut-parleurs que dans le plan horizontal incluant la tête de l'auditeur) donnent lieu à un sentiment de *présence* plus important que des systèmes de restitution sonore 3D (*i.e.* ayant des haut-parleurs partout autour de l'utilisateur). Le système le plus complexe et le plus fidèle d'un point de vue physique (donc le plus *immersif*) n'est donc pas celui qui procure le plus de *présence*.

- *L'authenticité et la plausibilité de l'environnement virtuel* : la cohérence globale de l'environnement virtuel, allant de la création d'un scénario plausible à la synthèse de réactions des personnages virtuels ou des objets virtuels qui soient en accord avec le monde virtuel, vont pousser le participant à croire au monde virtuel. Si le participant *n'y croit pas*, alors la sensation de *présence* en souffrira irrémédiablement.
- *La nature de la tâche ou de l'activité* : dans tout environnement virtuel, une part importante du contenu est la nature de la tâche ou de l'activité à laquelle participe l'utilisateur. La complexité d'exécution de la tâche ainsi que les possibilités d'interagir avec l'environnement virtuel qu'elle offre participent alors directement à la sensation de *présence*.

Ainsi, une interface destinée à des applications de réalité virtuelle devra chercher à être la plus *transparente* possible (description formelle de l'interface) mais devra aussi, et de façon complémentaire, créer un monde virtuel crédible accompagné d'une tâche stimulante (description contextuelle de l'interface). Ces deux aspects participant de façon égale au sentiment de *présence*, il est assez difficile de les étudier séparément l'un de l'autre. Dans tout ce qui suit, les efforts seront néanmoins portés sur la *transparence* de l'interface plutôt que sur le *contenu* proposé par l'interface.

2.3 Rendu sonore spatialisé et intérêt en environnement audio-visuel virtuel

Dans cette section, la façon dont est envisagé le concept de *rendu sonore spatialisé* est d'abord définie. La littérature concernant l'utilité, ou l'intérêt, d'un *rendu sonore spatialisé* pour des applications audio-visuelles de réalité virtuelle est ensuite détaillée. En particulier, son impact sur la *présence*, la *qualité perçue*, et la *compréhension de scènes complexes* est analysé. Cela permet de justifier en quoi l'ajout d'un *rendu sonore spatialisé* est bénéfique à un système *audio-visuel* destiné à la réalité virtuelle. Cet état de l'art sur le sujet permet aussi de cerner les questions encore sujettes à discussion.

2.3.1 Rendu sonore spatialisé

En accord avec la définition de la *réalité virtuelle* donnée à la section 2.2.1, un rendu sonore spatialisé devra *simuler dans un environnement audio virtuel le comportement d'entités autonomes qui sont en interaction en temps réel entre elles et avec un ou des utilisateurs en immersion pseudo-naturelle*. Un *rendu sonore spatialisé* aura donc pour objectif de restituer en temps réel, pour un ou plusieurs auditeurs, des sources sonores et l'environnement sonore qui leur est associé dans un espace virtuel tri-dimensionnel [Begault 1994, Tsingos & Warusfel 2006].

Pour créer un *rendu sonore spatialisé*, différents attributs des sources virtuelles doivent être restitués à l'auditeur. En premier lieu le système de rendu sonore devra permettre à l'auditeur de localiser les différentes sources virtuelles en termes de *distance* et de *direction* (voir figure 2.3.a). De plus, dans les environnements réels couramment rencontrés, l'auditeur reçoit non seulement le son provenant *directement* de la source sonore, mais aussi *indirectement* les sons résultants des réflexions, diffractions du son direct avec l'environnement réel ou des objets le composant. Le système de rendu sonore spatialisé devra donc

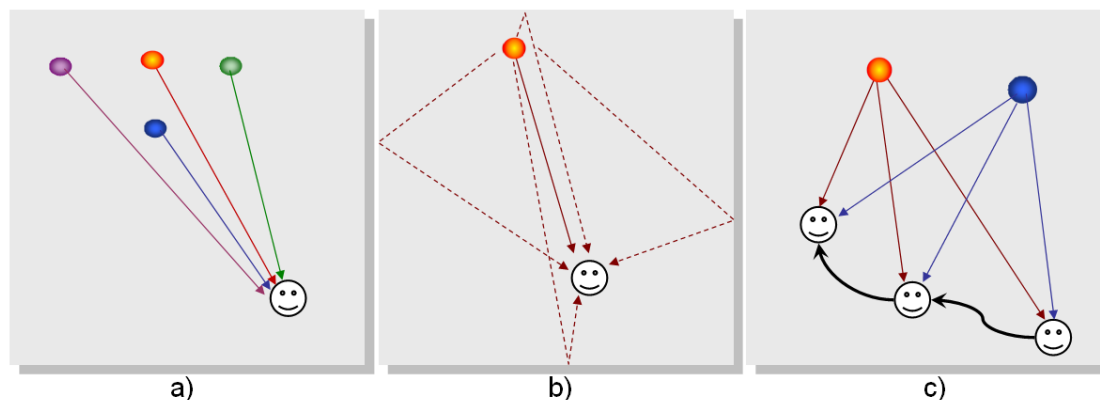


FIG. 2.3 – Un *rendu sonore spatialisé* idéal doit restituer aux auditeurs : a) la position des sources virtuelles (*distance, direction*), b) l'interaction des sources virtuelles avec leur environnement (*réverbération*), et c) les mouvements des sources virtuelles associés aux mouvements de l'auditeur (*effet de parallaxe*).

restituer aussi la *réverbération*, c'est à dire l'interaction des sources virtuelles avec l'environnement virtuel (voir figure 2.3.b). Enfin, lorsque l'auditeur se déplace dans la scène sonore qui lui est proposée, la perspective sonore perçue est modifiée et les sources sonores se déplacent relativement à lui. Cet effet est appelé *effet de parallaxe* et doit lui aussi être restitué à l'auditeur (voir figure 2.3.c) par le système de rendu sonore spatialisé.

Un *rendu sonore spatialisé* adapté à des applications de réalité virtuelle devra donc être capable de simuler les différents attributs mentionnés précédemment (*direction, distance, réverbération, effet de parallaxe*) pour maximiser l'immersion des auditeurs dans un *environnement audio virtuel*.

2.3.2 Impact sur la sensation de présence

L'ajout d'un *rendu sonore spatialisé* à un système de réalité virtuelle augmente a priori l'*immersion*. Mais, comme mentionné à la section 2.2.3, une augmentation de l'*immersion* n'est pas nécessairement corrélée à une augmentation de la *présence*. Il est donc légitime de se demander en premier lieu ce qu'un *rendu sonore* (indépendamment des attributs spatiaux qu'il simule) apporte à la sensation de *présence*. On montrera que l'*espace* est important pour la sensation de *présence*. Dans ces conditions, la question de l'impact des attributs spatiaux simulés par un *rendu sonore spatialisé* sur la *présence* perçue devient pertinente et sera étudiée.

La nécessité d'ajouter un *rendu sonore* dans les systèmes de réalité virtuelle a d'abord été déduite d'observations faites dans le monde réel. En effet, Gilkey & Weisenberger [1995] ont constaté que des anciens combattants de la seconde guerre mondiale ayant subi de graves lésions auditives ressentaient le monde réel comme «mort» et s'en sentaient «déconnectés». Cette constatation a été reprise par Murray et al. [2000] qui ont analysé les réponses à des questionnaires de participants ayant passé 20 minutes de leur vie quotidienne avec les conduits auditifs obstrués. Leur conclusion est que la modalité auditive est très importante dans le ressenti du monde environnant. Les études menées par Gilkey & Weisenberger [1995], Murray et al. [2000] suggèrent alors qu'un *rendu sonore* est nécessaire pour créer un sentiment de *présence* fort dans les applications de réalité virtuelle. Partant de ce constat, les études [Hendrix & Barfield 1995, 1996, Dinh

et al. 1999, Darken et al. 1999, Larsson et al. 2001] ont ensuite empiriquement montré que pour des participants plongés dans un monde virtuel, l'ajout d'un *rendu sonore* augmentait significativement la sensation de *présence*. L'augmentation de l'*immersion* créée par l'ajout d'un *rendu sonore* à un système de réalité virtuelle a donc pour conséquence directe l'augmentation de la sensation de *présence*.

L'ajout d'un *rendu sonore* aux systèmes de réalité virtuelle permet de stimuler la modalité auditive, mais en général ne procure pas une perception *spatiale* de l'environnement sonore. Witmer & Singer [1998], dans une étude impliquant 152 participants, établissent un questionnaire permettant de mesurer la *présence* dans lequel le «*naturel*» est identifié comme l'une des trois dimensions de la *présence*. Selon leurs termes, le «*naturel*» décrit «*dans quelle mesure [...] l'environnement virtuel est cohérent avec la réalité*». Dans une autre étude basée sur les réponses de 604 personnes à un questionnaire, Lessiter & Freeman [2001] identifient le «*sens de l'espace physique*» comme le facteur dominant dans la sensation de *présence*. De même, Schubert et al. [2001] trouvent dans les réponses de 296 sujets que la «*présence spatiale*» est la composante la plus importante de la sensation de *présence*. Plus spécifiquement, les résultats de l'étude menée par Ozawa et al. [2003] sur la *présence auditive* font aussi apparaître que la «*localisation sonore*» est un facteur déterminant dans la sensation de *présence*. Cela suggère alors qu'un *rendu sonore spatialisé* simulant certains ou tous les attributs spatiaux définis à la section 2.3.1 aura un impact positif sur la *présence* en comparaison d'un *rendu sonore* n'en simulant aucun.

Relativement peu d'études ont abordé la question de l'impact des attributs spatiaux d'un *rendu sonore spatialisé* sur la *présence* dans les applications de réalité virtuelle audio-visuelle [Larsson et al. 2005, Sanchez-Vives & Slater 2005]. Dans la première étude sur le sujet, Hendrix & Barfield [1995, 1996] ont montré que lors de la navigation dans un environnement virtuel, un *rendu sonore spatialisé* simulant *direction* et *distance* augmentait significativement la *présence* par rapport à un *rendu sonore spatialisé* ne simulant aucun de ces deux attributs (*i.e.* un rendu monophonique sur casque). Dans les études [Riecke et al. 2005, Valjamae 2005], l'impact d'un *rendu sonore spatialisé* simulant *distance* et *direction* comparé à un *rendu sonore spatialisé* ne simulant aucun de ces deux attributs est étudié sur la *vection*³ en environnement virtuel audio-visuel. Les résultats obtenus indiquent que l'ajout d'un *rendu sonore spatialisé* simulant *distance* et *direction* augmente légèrement mais significativement la sensation de *présence*. Lessiter & Freeman [2001], Freeman & Lessiter [2001] ont étudié comment l'ouverture angulaire sur laquelle la *direction* est rendue impacte la *présence*. Leurs résultats indiquent qu'un rendu des *directions* frontales augmente la sensation de *présence*. Bormann [2005] a étudié l'impact sur la *présence* du rendu de la *distance* par un *rendu sonore spatialisé*. Des sujets devaient rechercher une source sonore active associée à un objet visuel dans un environnement virtuel abstrait. Le *rendu sonore spatialisé* dont ils disposaient était soit complet (première condition : *direction* et *distance* de la source sonore, voir 2.3.1), soit partiel (seconde condition : seulement la *direction* de la source sonore). Étonnamment, plus de *présence* a été ressentie dans la seconde condition. Dans les travaux d'Ozawa & Miyasaka [2004], l'effet de la *distance* sur la *présence* est étudiée dans un contexte audio-visuel en faisant varier le niveau sonore global. Les résultats obtenus indiquent que la *présence* est maximale lorsque la *distance* audio restituée est cohérente avec la *distance* visuelle.

3. La *vection* est une sensation de mouvement induite par un stimulus extérieur, audio ou visuel. Ce mouvement est ressenti par exemple lorsque, pour une personne assise dans un train stationné en gare, le train placé sur le quai d'en face démarre et lui donne l'impression que c'est son train qui démarre.

De façon similaire, Larsson [2002], Larsson et al. [2007] ont étudié l'influence de la *réverbération* sur la sensation de *présence*. Dans [Larsson et al. 2007], les sujets écoutent un concert d'orgue enregistré en champ proche dans une église virtuelle au sein de laquelle l'interaction des sources sonores avec leur environnement est cohérente ou non (pas de *réverbération*, *réverbération* cohérente, *réverbération* d'une salle quatre fois plus petite). Les résultats obtenus ne montrent cependant pas de différences statistiquement significatives sur la *présence* perçue entre les différentes conditions. Cependant, il semblerait que la condition où la *réverbération* était cohérente avec le rendu visuel ait augmenté la *présence*.

L'addition d'un *rendu sonore spatialisé* capable de simuler *direction*, *distance* et *réverbération* à une application de réalité virtuelle audio-visuelle augmente donc généralement la *présence* ressentie par les utilisateurs. Cependant, le rôle de chacun des attributs du *rendu sonore spatialisé* (voir section 2.3.1) dans cette augmentation reste encore largement à déterminer, en particulier en ce qui concerne l'*effet de paralaxe*.

2.3.3 Impact sur les réactions émotionnelles

Les systèmes audio-visuels cherchent parfois à véhiculer des *émotions* jusqu'à leurs utilisateurs. Au cinéma par exemple, il n'est pas rare d'entendre un extrait musical que l'on pourrait qualifier de «*triste*» pour accompagner une scène de séparation. Les applications de réalité virtuelle audio-visuelles peuvent aussi chercher à induire certaines *réactions émotionnelles*. Cependant, la partie non-informative des signaux auditifs (*i.e.* tous les attributs d'un son qui ne lui donnent pas un *sens* particulier) reste largement inexploitée à cet effet mais est pourtant très importante [Tajadura-Jimenez & Vastfjall 2008]. Par exemple, l'étude menée par Vastfjall et al. [2002] a cherché à déterminer dans quelle mesure dans un *rendu sonore spatialisé*, la *réverbération* (voir section 2.3.1), pouvait affecter les *réactions émotionnelles* d'utilisateurs par rapport à un objet virtuel. Dans ce but, 76 participants ont dû évaluer les émotions qu'ils ressentaient (sur une échelle avec deux dimensions : *plaisir* et *excitation*) à l'écoute de pièces musicales rendues avec différents niveaux de *réverbérations* dans un environnement virtuel audio-visuel. Les résultats obtenus indiquent que la *réverbération* influence les *réactions émotionnelles* des sujets en diminuant leur sensation de «*plaisir*». Dans une autre étude, Vastfjall [2003] s'est intéressé à l'impact d'un *rendu sonore spatialisé* simulat ou non la *direction* des sources sonores sur les *réactions émotionnelles* de participants et ce dans un environnement virtuel uniquement audio. Les résultats obtenus indiquent que les systèmes de *rendu sonore spatialisé* simulat la *direction* des sources sonores induisent des changements des *réactions émotionnelles* significativement plus forts ceux ne la simulat pas.

Les différents attributs d'un *rendu sonore spatialisé* semblent donc avoir un impact non négligeable sur les *réactions émotionnelles* de participants plongés dans un environnement audio-visuel virtuel. Il existe cependant très peu d'études sur le sujet et ces résultats mériteraient donc d'être approfondis. Par ailleurs, il a été constaté que les changements observés sur les *réactions émotionnelles* sont corrélés à des changements observés sur la sensation de *présence* [Vastfjall 2003]. Il existe donc peut-être un lien de causalité circulaire entre *réactions émotionnelles* et *présence*, l'émotion favorisant la présence et la présence renforçant l'émotion [Bouvier 2008]. Les résultats reportés dans cette section sont donc à rapprocher de ceux décrits à la section 2.3.2.

2.3.4 Impact sur la compréhension d'environnements complexes

Alors que la vision constitue un sens très directionnel, l'audition est quant à elle un sens plutôt omnidirectionnel. En effet, on ne *voit* que dans une direction donnée de l'espace alors qu'on *entend* dans toutes les directions simultanément (voir section 7.3.1). De ce fait, certains bénéfices des *rendus sonores spatialisés* dans des applications de réalité virtuelle peuvent être attendus de cette complémentarité spatiale entre l'ouïe et la vue, en particulier en ce qui concerne la compréhension et la navigation dans des environnements audio-visuels complexes.

L'*intelligibilité* d'une scène audio-visuelle complexe est la faculté de comprendre, d'interpréter, un flux audio-visuel parmi les différents flux simultanés composant la scène. La question de l'apport des différents attributs d'un *rendu sonore spatialisé* à l'*intelligibilité* a d'abord été étudiée dans les environnements audio uniquement. Par exemple, Hawley et al. [1999] ont étudié l'influence d'un *rendu sonore spatialisé* simulant la *direction* de provenance des sources sonores sur l'*intelligibilité* de la parole lorsque plusieurs personnes parlent simultanément. Leurs résultats montrent que la simulation de la *direction* est en général bénéfique à l'*intelligibilité* de la parole. Sanson & Warusfel [2010] ont étudié l'influence de la précision du rendu de la *direction* sur la capacité de sujets à détecter des changements dans l'organisation spatiale d'une scène sonore faisant intervenir trois personnes simultanément. Ils montrent ainsi que le système de *rendu sonore spatialisé* restituant le plus fidèlement la *direction* des personnes permet une meilleure discrimination des changements. Saupe et al. [2010] se sont quand à eux intéressés à l'*intelligibilité* de signaux musicaux. Ils ont étudié la faculté d'un auditeur à distinguer de larges intervalles dans la mélodie d'une pièce polyphonique en fonction de la *direction* de provenance de chacune des voix. Leurs résultats montrent qu'une séparation spatiale des différentes voix permet une meilleure détection des larges intervalles. Des résultats similaires ont été obtenus dans un contexte de contrôle du trafic aérien ou les participants étaient en plus divertis par une tâche visuelle [MacDonald et al. 2002]. Dans une analyse bibliographique, Bronkhorst [2000] étudie les résultats présentant l'influence de la *réverbération* sur l'*intelligibilité*. Il mentionne le fait que dans la *réverbération*, les premières réflexions augmentent l'*intelligibilité* alors que la réverbération tardive la diminue. Dans un environnement audio virtuel, la simulation des différents attributs d'un *rendu sonore spatialisé* permet donc de focaliser l'attention de l'auditeur vers une zone spatiale précise de la scène virtuelle ce qui a pour conséquence de rendre la scène proposée plus *intelligible* et plus facile à analyser.

L'impact d'un *rendu sonore spatialisé* sur l'*intelligibilité* a aussi été étudié dans des environnements audio-visuels complexes, particulièrement pour des applications de téléconférence. Walker et al. [2000] ont par exemple montré qu'avec 5 personnes parlant simultanément, une conversation précise n'était *intelligible* que dans la condition faisant intervenir un *rendu sonore spatialisé* simulant la *direction* des sources sonores. De façon analogue, de Bruijn & Boone [2003], Boone [2003] ont mis en évidence qu'un *rendu sonore spatialisé* capable de restituer la *distance* permettait d'augmenter l'*intelligibilité* de la parole. Baldis [2001] a quant à elle démontré expérimentalement que l'ajout de la *direction* dans un *rendu sonore spatialisé* permet d'augmenter significativement la compréhension et la mémorisation de conversations entendues dans le cadre de téléconférences. Ces résultats ont été partiellement dupliqués par Kilgore et al. [2003], Kilgore [2009] qui ont de plus spécifié que les bénéfices augmentaient avec le nombre de voix parlant simultanément. Dans les environnements audio-visuels recréant des situations de téléconférences,

les différents attributs d'un *rendu sonore spatialisé* s'avèrent donc très utiles en ce qui concerne la compréhension globale des scènes présentées.

Comme la *réalité virtuelle* vise à faire des participants des *acteurs* au sein du monde virtuel et non de simples *spectateurs* (voir section 2.2.1), il est important de savoir aussi comment les différents attributs d'un *rendu sonore spatialisé* influencent la capacité des utilisateurs à savoir où ils sont et où ils vont dans un environnement virtuel complexe. La *navigation* est le processus permettant aux utilisateurs de contrôler leurs mouvements en utilisant des indices issus de l'environnement virtuel de telle sorte qu'ils puissent atteindre leurs objectifs sans s'égarer [Darken & Sibert 1993]. Dans un environnement virtuel audio uniquement, [Loomis et al. 1998, Lokki et al. 2000] ont montré que la *navigation* est possible quand les attributs de *direction*, de *distance* et l'*effet de parallaxe* sont disponibles. La *réverbération* semble cependant dégrader les performances de *navigation* des utilisateurs [Lokki et al. 2000]. Dans un environnement virtuel audio-visuel, Darken & Sibert [1993] suggèrent qu'un *rendu sonore spatialisé* restituant la *direction* peut permettre d'estimer grossièrement la direction d'une cible et donc devrait améliorer les possibilités de *navigation* des utilisateurs. Ce résultat est rigoureusement établi par Gröhn et al. [2005] qui montrent que dans un environnement audio-visuel virtuel où le *rendu sonore spatialisé* restitue la *direction*, la *distance* et l'*effet de parallaxe*, la «*navigation en utilisant les indices audio-visuels est remarquablement plus facile et rapide qu'en utilisant seulement les indices auditifs ou visuels*». Les indices auditifs sont utilisés au début pour localiser la cible et guider le début de la navigation. Les indices visuels sont utilisés en fin de parcours pour arriver précisément à la cible. Gröhn [2006] montre aussi que la *direction* simulée par un *rendu sonore spatialisé* peut aider à l'orientation dans le monde virtuel. Les différents attributs d'un *rendu sonore spatialisé*, et en particulier l'*effet de parallaxe*, rendent donc la *navigation* dans les espaces virtuels audio-visuels plus facile.

Les *rendus sonores spatialisés* apportent donc des bénéfices non négligeables dans des applications audio-visuelles de réalité virtuelle où la compréhension et la navigation dans des environnements audio-visuels complexes sont importantes.

2.3.5 Impact sur la *qualité globale audio-visuelle perçue*

D'un point de vue pratique, la notion de *qualité* est très importante pour les applications audio-visuelles ludiques ou dédiées au divertissement. Une définition de cette notion est rapportée dans Pellegrini [2001] comme étant le «*résultat de l'appréciation perceptive de la nature d'une entité par rapport aux attentes concernant cette entité*». La notion de *qualité* est ainsi à la fois subjective et circonstancielle et fait intervenir explicitement le contenu audio-visuel.

Cependant, les *qualités* audio et visuelle perçues ne sont pas indépendantes l'une de l'autre. Cela a été noté de longue date par le monde de l'industrie audio-visuelle qui estime en plus que la *qualité audio-visuelle globale* est liée à d'autres notions comme la *présence* (voir sections 2.2.2 et 2.3.2) ou l'*intelligibilité* (voir section 2.3.4) [Woszczyk et al. 1995]. Cette interdépendance a été expérimentalement démontrée [Storms 1998, Beerends & de Caluwe 1999, Hollier et al. 1999, Storms & Zyda 2000]. Dans une synthèse bibliographique sur ce sujet, You et al. [2010] concluent que «*les qualités audio et visuelle contribuent toutes les deux à la qualité globale et que leur produit est fortement corrélé à la qualité globale*». D'autre part, ils montrent que dans certaines situations la qualité audio peut être déterminante de la qualité

globale perçue (téléconférence, clips musicaux). Dans le cas d'applications de type «*téléconférence*» cela est effectivement rapporté par Walker et al. [2000]. Dans une autre synthèse bibliographique sur ce sujet, Kohlrausch & van de Par [2005] expliquent que certains auteurs estiment même qu'une «*dégradation audio est plus nuisible à la qualité globale qu'une dégradation visuelle*». Dans un système de rendu audio-visuel, augmenter la qualité d'un *rendu sonore* est donc potentiellement très intéressant car cela peut contribuer à l'augmentation de la *qualité globale perçue*.

L'impact d'un *rendu sonore spatialisé* sur la *qualité globale audio-visuelle perçue* reste cependant assez flou. La sensation d'espace audio peut aider à augmenter la *qualité globale audio-visuelle perçue* comme le montrent Bech et al. [1995], mais très peu d'études ont porté sur ce sujet. Freeman & Lessiter [2001], Lessiter & Freeman [2001] ont par exemple étudié comment l'ajout d'un *rendu sonore spatialisé* restituant la *direction* influait sur la *qualité globale perçue*. Ils ne mettent cependant en évidence aucun effet significatif de l'ajout de la *direction* sur la *qualité globale perçue*. D'autre part, ils constatent une augmentation significative seulement sur certaines dimensions de l'expérience audio (*naturel de la scène* et *investissement personnel*). Bormann [2008] montre expérimentalement que dans le cas où un *rendu sonore spatialisé* est présent, la qualité perçue du rendu visuel est très influencée par l'importance de la modalité auditive dans la tâche effectuée. Même si les enjeux commerciaux semblent considérables, très peu d'attention est aujourd'hui portée sur l'éventuel impact des différents attributs d'un *rendu sonore spatialisé* sur la *qualité audio-visuelle globale*.

Les conséquences, en terme de *qualité audio-visuelle*, d'incohérences spatiales entre le rendu visuel et le *rendu sonore spatialisé* ont été plus étudiées. Komiyama [1989] montre qu'un écart en *direction* pouvant aller jusqu'à 20° en azimut est «*acceptable*» dans le contexte de la télévision haute-définition. de Bruijn et al. [2000] suggèrent que des écarts en *direction* plus grands sont susceptibles d'être tolérés dans le plan vertical. En terme d'*effet de parallaxe*, de Bruijn & Boone [2002], Melchior et al. [2003], Sporer & Klehs [2004] montrent expérimentalement que des incohérences entre les rendus audio et visuels sont perçues comme *gênantes*. Ainsi, assurer la cohérence entre le rendu visuel et le *rendu sonore spatialisé*, au moins en ce qui concerne la *direction* et l'*effet de parallaxe*, est un bon moyen d'augmenter la *qualité audio-visuelle globale*.

L'ajout d'un *rendu sonore spatialisé* à un dispositif audio-visuel pour la réalité virtuelle peut donc s'avérer bénéfique à la *qualité audio-visuelle globale perçue* par les utilisateurs, sous réserve que le rendu visuel et le *rendu sonore spatialisé* soient cohérents entre eux.

2.3.6 Bilan

L'objet de cette partie était d'abord de définir le concept de *rendu sonore spatialisé* (voir section 2.3.1). Dans l'état actuel de la science, un *rendu sonore spatialisé* est ainsi caractérisé par quatre attributs : la *direction*, la *distance*, la *réverbération* et l'*effet de parallaxe*. La littérature concernant l'impact de chacun de ces attributs pour des applications audio-visuelles de réalité virtuelle a ensuite été détaillée. En particulier, les impacts des différents attributs sur la *présence* (section 2.3.2), les *réactions émotionnelles* (section 2.3.3), la *compréhension de scènes complexes* (section 2.3.4) et la *qualité perçue* (section 2.3.5) ont été analysés. Une synthèse sous forme de tableau de cette étude bibliographique est proposée dans le tableau 2.1.

De cette synthèse bibliographique, il ressort deux points principaux :

2.4. Intégration d'un système de rendu sonore spatialisé au sein d'une interface visuelle de réalité virtuelle typique

	Présence perçue	Émotions	Intelligibilité	Performances de navigation	Qualité audio-visuelle
Direction	Augmentée	Intensité des réactions augmentée	Augmentée	Augmentées	Augmentée en général et diminuée en cas d'incohérences AV
Distance	Diminuée, et maximale en cas de cohérence AV	?	Augmentée	Augmentées	?
Effet de parallaxe	?	?	Augmentée en théorie	Augmentées	Diminuée en cas d'incohérences AV
Réverbération	Augmentée	Sensation de plaisir diminuée	Augmentée par les réflexions précoces et diminuée par la réverbération tardive	Diminuées	?

TAB. 2.1 – Résumé de l'impact des différents attributs d'un *rendu sonore spatialisé* (*direction*, *distance*, *effet de parallaxe* et *réverbération*) sur différents aspects de la perception d'un environnement audio-visuel de réalité virtuelle.

- L'ajout d'un *rendu sonore spatialisé* a un **impact considérable** sur un grand nombre de facteurs qui semblent importants dans diverses applications audio-visuelles de réalité virtuelle (présence, émotions, intelligibilité, performances de navigation, qualité audio-visuelle globale, etc.). La **cohérence audio-visuelle spatiale** semble jouer un rôle important dans cet impact.
- L'étude de l'impact des différents attributs d'un *rendu sonore spatialisé* sur ces différents facteurs est encore **incomplète**. Il reste donc beaucoup à faire pour conforter certains résultats ou tout simplement obtenir de nouveaux résultats dans ce domaine. En particulier, le fait que la *distance* n'augmente pas la sensation de *présence* semble assez curieux.

Au vu de cette étude, il est justifié de chercher à ajouter un *rendu sonore spatialisé* à un système *audio-visuel* destiné à la réalité virtuelle. La mise au point d'un tel type de système peut permettre non seulement d'améliorer considérablement les performances des applications de réalité virtuelle, mais aussi d'étudier les impacts de différents attributs du son spatialisé sur ces performances.

2.4 Intégration d'un système de rendu sonore spatialisé au sein d'une interface visuelle de réalité virtuelle typique

Dans cette section, les compromis techniques liés à l'ajout d'un *rendu sonore spatialisé* dans une interface visuelle de réalité virtuelle typique sont abordés. Le problème à résoudre peut être formulé de cette façon : *comment, dans une interface visuelle typique, véhiculer jusqu'aux deux oreilles d'un utilisateur des stimuli audio qui soient cohérents avec sa position et ses déplacements dans le monde virtuel ?* Une interface visuelle de réalité virtuelle classique, la CAVE (*Cave Automatic Virtual Environments*), est introduite à la section 2.4.1. Puis, les technologies audio permettant la création d'un *rendu sonore spatialisé*

sont brièvement présentées et comparées (voir section 2.4.2). Enfin les compromis à effectuer lors de leur intégration avec cette interface visuelle de réalité virtuelle typique sont mis en avant et illustrés sur des systèmes existants (voir section 2.4.3).

2.4.1 Une interface visuelle de réalité virtuelle typique : la CAVE

De nombreuses interfaces visuelles pour la réalité virtuelle existent actuellement. Dans cette section on se limitera cependant à la description d'une interface visuelle fixe et permettant la reproduction d'un environnement virtuel à taille humaine. Ce type d'interface est communément appelée : «*interface à projection sur grands écrans*». Pour de plus amples détails concernant les autres interfaces visuelles existantes, on se référera à la synthèse présentée récemment par Fuchs & Dominjon [2006].

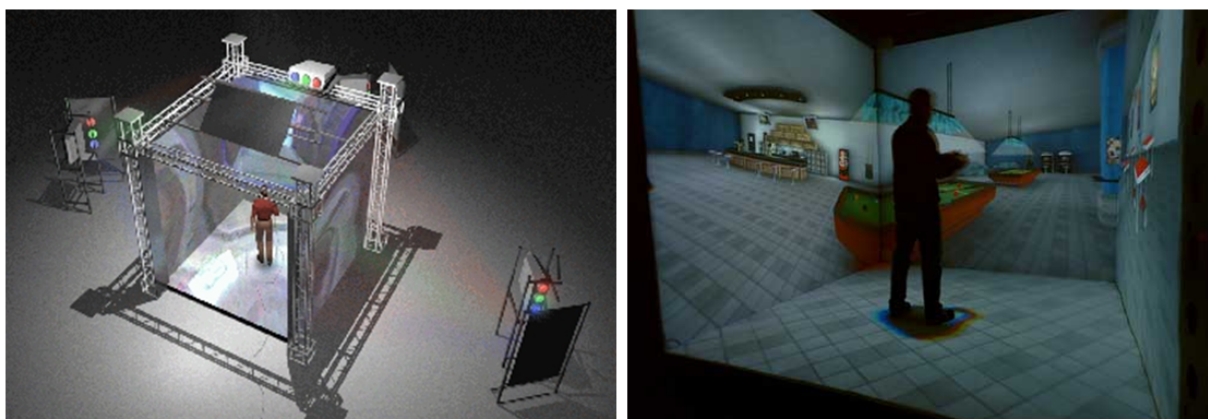


FIG. 2.4 – Un exemple de système de type CAVE : le SAS Cube créé par la société CLARTE. Sur l'image de gauche, une vue d'ensemble du système est proposée : Vidéo-projecteurs, miroirs et écrans. L'image de droite montre un utilisateur interagissant avec le monde virtuel dans lequel il est immergé, images issues du site : <http://www.clarte.asso.fr>

Pour procurer une bonne sensation de *présence* à un ou plusieurs utilisateurs, une interface visuelle pour la réalité virtuelle devra de manière générale chercher à offrir à l'utilisateur un grand champ de vision horizontal et vertical tout en affichant les images avec une résolution suffisante par rapport à l'acuité humaine (voir section 7.2.2). Les interfaces visuelles «*à projection sur grands écrans*» reposent ainsi sur un ou plusieurs écrans plats qui sont couplés à un certain nombre de vidéo-projecteurs. La CAVE (*Cave Automatic Virtual Environments*) développée par Cruz-Neira et al. [1992, 1993] est le premier représentant de cette classe d'interfaces visuelles. Cette interface est composée de plusieurs écrans (de 2 à 6) disposés de façon à former tout ou une partie d'un cube (voir par exemple le SAS Cube sur la figure 2.4). Des images sont alors projetées sur ces écrans, ce qui procure à l'utilisateur un rendu visuel avec un grand champ de vision et un monde virtuel dans lequel il peut évoluer.

Les systèmes de type CAVE sont ainsi des systèmes fixes permettant une reproduction d'un monde virtuel à taille réelle les plus répandus et les mieux maîtrisés techniquement [Fuchs & Dominjon 2006]. Comparativement aux autres interfaces visuelles classiquement utilisées pour la réalité virtuelle, les CAVEs permettent de maximiser le champ de vision, sont des systèmes très peu intrusifs et peuvent accueillir

simultanément plusieurs utilisateurs [Cruz-Neira et al. 1992]. L'intégration des technologies de création de *rendu sonore spatialisé* dans ce type de systèmes est donc une question technologique importante.

2.4.2 Technologies de création de rendu sonore spatialisé

Pour acheminer les signaux audio jusqu'aux oreilles de l'utilisateur et créer un *rendu sonore spatialisé* au sein d'interfaces visuelles de type CAVE, deux solutions sont possibles [Tsingos & Warusfel 2006] : l'utilisation de *casques audio* (système porté) ou bien l'utilisation d'un *réseau de haut-parleurs* correctement pilotés (système externe).

Utilisation de *casques audio*

L'utilisation de *casques audio* est avantageuse car elle permet d'envoyer directement les signaux voulus aux oreilles de l'utilisateur. Elle pose cependant le problème de la détermination des signaux à envoyer en fonction de la position de l'utilisateur dans l'espace virtuel. On note ici que pour pouvoir utiliser un rendu sur casque de façon dynamique, la position de l'utilisateur doit être connue à chaque instant. Cela implique donc l'utilisation conjointe d'un système de suivi de l'utilisateur (système de *tracking*). Plusieurs approches sont alors possibles pour déterminer les signaux à envoyer à chaque oreille :

- *La stéréophonie* [Tsingos & Warusfel 2006] : c'est la technologie de rendu spatialisé la plus facile à mettre en œuvre. Elle est basée sur une illusion perceptive qui permet la création d'une *source fantôme* positionnable entre deux sources réelles. Cette technologie est cependant peu utilisée car elle donne à l'utilisateur l'impression que les sources sonores sont localisées à l'intérieur de sa tête (problème d'*externalisation*). Cette solution est donc mal adaptée à la création d'un *rendu sonore spatialisé* basé sur des casques audio.
- *Le rendu binaural* [Begault 1994, Tsingos & Warusfel 2006, Nicol 2010] : c'est la technologie de rendu sonore spatialisé sur casque la plus répandue. Elle permet de pallier partiellement le problème d'*externalisation* en réalisant numériquement un filtrage du son émis par la source sonore positionné dans l'espace équivalent à celui réalisé par les pavillons, la tête, le torse et les épaules de l'utilisateur. Ces filtres sont nommés «*Head Related Transfert Functions*» (HRTFs). Cependant, pour que la sensation d'*externalisation* soit bien réelle, il faut que les HRTFs soient adaptées à chaque individu (problème d'*individualisation*) ou que les utilisateurs aient appris à écouter avec des HRTFs qui ne sont pas les leurs [Nicol 2010]. Cette solution permet donc la création d'un *rendu sonore spatialisé* satisfaisant sous réserve de disposer d'HRTFs *individualisées* et d'un système de *tracking*.

Utilisation d'un *réseau de haut-parleurs*

L'utilisation d'un *réseau de haut-parleurs* est une alternative avantageuse aux *casques audio* car elle permet d'éviter d'équiper l'utilisateur d'un casque. Son principal désavantage réside dans le fait que l'on ne contrôle pas directement les signaux atteignant les oreilles de l'utilisateur, mais uniquement ceux qui sont émis par les haut-parleurs. Il faut donc être en mesure de déterminer quels signaux envoyer aux haut-parleurs et plusieurs techniques sont disponibles à cet effet :

- *Les techniques basées sur la stéréophonie* [Tsingos & Warusfel 2006] : comme on l'a vu précédemment, la stéréophonie sur casque ne crée pas de sensation d'*externalisation* des sources sonores. L'utilisation d'un réseau de haut-parleurs pilotés par de la stéréophonie ou par ses dérivées 2D (systèmes multicanaux : 5.1, 7.1, etc.) ou 3D («*Vector Based Amplitude Panning*», VBAP, voir [Pulkki 1997]) permet d'éviter ce problème. Cependant, la synthèse de l'espace offerte par ces technologies n'est correcte que dans une toute petite zone de l'espace, le «*sweet spot*⁴», ce qui implique des utilisations où les participants sont fixes et rassemblés autour d'un unique point d'écoute. Cette technologie ne permet donc pas la création d'un rendu sonore spatialisé dynamique.
- *Binaural sur haut-parleurs* [Moller 1989] : pour s'affranchir des limitations inhérentes au «*sweet spot*» il est possible de chercher à faire correspondre à chaque instant la zone de rendu optimal avec la position de l'utilisateur en utilisant des techniques de traitement du signal. Cela nécessite donc d'utiliser aussi un système de *tracking*. Ces techniques, dites *binaurales sur haut-parleurs*, doivent ensuite être couplées à des systèmes de synthèse binaurale ce qui fait apparaître à nouveau les problèmes d'*individualisation* évoqués précédemment. Dans le cas où la position de l'utilisateur n'est pas suivie en temps réel, cette technique est dite *transaurale* et est extrêmement sensible à la position de l'utilisateur (*sweet spot*).
- *L'holophonie* : une autre façon de s'affranchir du «*sweet spot*» est de chercher à reproduire physiquement un champ sonore dans une grande zone de l'espace. Ces technologies, dites *holophoniques* peuvent être divisées en deux catégories : la synthèse *ambisonique* [Gerzon 1985, Daniel 2001] et la «*Wave Field Synthesis*» (WFS) [Berkhout et al. 1993, Corteel 2004, de Vries 2009]. La synthèse *ambisonique* permet la reconstruction d'un champ sonore dans un volume délimité par un réseau tridimensionnel de haut-parleurs à partir de sa décomposition en harmoniques sphériques. La taille de la zone d'écoute obtenue dépend de la taille du réseau de haut-parleurs et diminue avec la fréquence. La synthèse *ambisonique* permet donc la création d'un rendu sonore spatialisé 3D dans une zone de rendu dont la taille diminue avec la fréquence. La WFS permet la reconstruction d'un champ sonore quelconque dans un volume délimité par un réseau bi-dimensionnel de haut-parleurs (voir section 2.5.3 pour plus de détails). La taille de la zone d'écoute obtenue est, contrairement à la synthèse *ambisonique*, stable en fréquence mais ce jusqu'à une certaine fréquence où le champ sonore n'est plus correctement reconstruit (fréquence d'*aliasing*, voir section 2.5.3). La WFS permet donc la création d'un rendu sonore spatialisé 2D dans une assez grande zone de rendu.

Les possibilités et contraintes proposées par les différentes technologies disponibles sont résumées dans le tableau 2.2. Ces possibilités et contraintes sont très importantes dans la phase de conception d'un système destiné à la réalité virtuelle car elles auront un impact direct sur la qualité du *rendu sonore spatialisé* et sur les applications possibles du système.

2.4.3 Choix d'une technologie pour l'intégration d'un *rendu sonore spatialisé* avec une interface visuelle de type CAVE

Il a été montré à la section 2.3.6 que l'ajout d'un *rendu sonore spatialisé* à un système de réalité virtuelle peut avoir un impact considérable sur un nombre importants de facteurs dans diverses applications audio-

4. Plus exactement, il s'agit d'un problème de proximité aux haut-parleurs pour la VBAP.

2.4. Intégration d'un système de rendu sonore spatialisé au sein d'une interface visuelle de réalité virtuelle typique

	Rendu sur casque		Rendu sur réseau de haut-parleurs (HPs)				
	<i>Stéréophonie</i>	<i>HRTF</i>	<i>Multicanal</i>	<i>VBAP</i>	<i>Binaural sur HPs</i>	<i>Ambisonique</i>	<i>WFS</i>
Zone de rendu	Mobile	Mobile	Petite et fixe	Petite et fixe	Petite et mobile	Grande, fixe et qui diminue avec la fréquence	Grande et fixe, jusqu'à une certaine fréquence
Objets virtuels	Dans la tête	Partout	Entre les HPs	Entre les HPs	Partout	Derrière les HPs	Partout
Individualisation	Externalisation impossible	Nécessaire	Inutile	Inutile	Nécessaire	Inutile	Inutile
Suivi de la tête	Nécessaire	Nécessaire	Inutile	Inutile	Nécessaire	Inutile	Inutile
Utilisateurs	Multi	Multi	Mono	Mono	Mono	Multi	Multi
Dimensionnalité	2D	3D	2D	3D	3D	3D	2D
Calculs	Peu	Moyen	Peu	Peu	Moyen	Moyen	Beaucoup
Matériel	Casque	Casque	≥ 2 HPs	≥ 10 HPs	2 ou 4 HPs	≥ 8 HPs	≥ 12 HPs

TAB. 2.2 – Résumé des possibilités et des contraintes proposées par les différentes technologies dédiées à la création d'un *rendu sonore spatialisé*.

visuelles (présence, émotions, intelligibilité, performances de navigation, qualité audio-visuelle globale, etc.). La question est maintenant de savoir comment intégrer un système permettant la création d'un *rendu sonore spatialisé* dans une interface de type CAVE et quelle technologie choisir pour ce faire.

Systèmes portés ou systèmes externes ?

Les systèmes *portés* présentent l'avantage de s'intégrer facilement dans tout type d'interface visuelle de réalité virtuelle car ils nécessitent seulement d'équiper l'utilisateur d'un casque. En contrepartie, les casques audio constituent une technologie fortement intrusive, ce qui risque de limiter l'*immersion* et donc la sensation de *présence* qu'auront les utilisateurs du monde virtuel (voir section 2.2.3). D'autre part, pour d'éventuelles applications multi-utilisateurs, il est possible d'équiper chaque participant d'un casque, mais cela risque de rendre la communication verbale entre les utilisateurs, et donc les interactions, difficile [Hiipakka et al. 2001]. Les systèmes portés sont donc une solution de bonne qualité et aisée à mettre en œuvre, mais risquent de détériorer la sensation de *présence* et les possibilités d'*interactions* des utilisateurs entre eux, primordiales dans les applications de réalité virtuelle.

Comparés aux systèmes *portés*, les systèmes *externes* ont l'avantage de ne pas poser de problème d'individualisation et donc a priori de produire des *rendus sonores spatialisés* de meilleure qualité pour le plus grand nombre. De plus, les applications impliquant plusieurs utilisateurs sont relativement faciles à mettre en place, ne demandent pas de matériel supplémentaire, et autorisent une communication directe entre participants. Néanmoins, l'intégration des systèmes *externes* dans les dispositifs de réalité virtuelle impose de nouvelles contraintes pratiques auxquelles il faut trouver des solutions. En premier lieu, il faut être capable d'intégrer et de masquer efficacement des haut-parleurs dans l'interface visuelle de réalité virtuelle. D'autre part, les sons venant du monde réel sont susceptibles de déranger l'utilisateur plus que s'il portait un casque audio. La salle dans laquelle un tel système est installé doit donc faire l'objet d'un traitement acoustique spécifique pour être bien isolée du monde extérieur. Sous réserve qu'il soit possible de répondre

à ces contraintes, les systèmes *externes* garantissent alors potentiellement un rendu sonore spatialisé de très bonne qualité ne nuisant ni à la *présence*, ni à l'*interaction*.

Quelle technologie choisir pour les systèmes *externes* ?

Pour les systèmes *portés*, le choix de la technologie de rendu sonore est aisé étant donné qu'il est très limité. Dans la mesure du possible, il est ainsi préférable de choisir une technologie de rendu binaural basée sur des HRTFs pour piloter un système de rendu *porté* [de Sorbier de Pognadoresse et al. 2009].

Pour les systèmes *externes*, le choix de technologies permettant la création d'un *rendu sonore spatialisé* dans une interface visuelle est assez large. Comme le montre le tableau 2.2, différents choix sont possibles en fonction des applications visées :

1. Les technologies *stéréophoniques* basées sur l'illusion d'une source *fantôme*, *i.e.* les systèmes multicanaux et la VBAP proposent une restitution spatialement cohérente dans une très petite zone de l'espace. Cela interdit tout mouvement des participants sous peine de perte de l'impression d'espace sonore⁵. Certains systèmes audio-visuels de réalité virtuelle utilisent des systèmes multicanaux [Valbom & Marcos 2005, Bouvier et al. 2006] ou VBAP [Hiipakka et al. 2000, 2001, Naef et al. 2002, Ogi et al. 2003].
2. Les technologies *binaurales sur haut-parleurs* permettent de s'affranchir du *sweet spot*, mais au prix d'un rendu mono-utilisateur et d'éventuels problèmes d'individualisation. Le système proposé par Kuhlen et al. [2007], Lentz et al. [2007] utilise cette technologie.
3. Les techniques *holophoniques* constituent le meilleur compromis entre la taille de la zone d'écoute et la qualité de la restitution spatiale. Elles sont de plus naturellement adaptées à une écoute collective ou à une écoute dynamique des scènes sonores sans avoir besoin de système de *tracking*. Cependant, assez peu de systèmes audio-visuels dédiés à la réalité virtuelle utilisent l'ambisonique [Faria et al. 2005a,b, Amatriain et al. 2007, Thompson et al. 2009, Okamoto et al. 2010] et encore moins sont basés sur la WFS [Okubo et al. 2005, Springer et al. 2006, de Vries 2009, Altinsoy et al. 2010]

Comment intégrer une technologie de *rendu sonore spatialisé* par *holophonie* et basée sur un système *externe* avec l'interface visuelle ?

En pratique, l'intégration d'un système *externe* et *holophonique* permettant de créer un *rendu sonore spatialisé* dans une interface visuelle pour la réalité virtuelle pose différents problèmes.

Le premier problème réside dans le fait qu'en utilisant une interface visuelle comme celle présentée à la figure 2.4, il n'est pas possible de placer les haut-parleurs n'importe où. Il est en effet impossible de placer les haut-parleurs devant les écrans ou entre les vidéo-projecteurs et les écrans, sous peine de masquer le rendu visuel. La figure 2.5 présente quelques exemples d'intégration de systèmes *externes* avec des interfaces visuelles de type CAVE. Dans ce contexte, il y a donc des conflits de positionnement entre écrans, vidéo-projecteurs et haut-parleurs qui sont assez gênants. Or les technologies *holophoniques* ont des prérequis assez forts concernant le positionnement des haut-parleurs [Amatriain et al. 2007]. L'ambisonique suppose par exemple une répartition de haut-parleurs qui soit la plus homogène possible sur une sphère.

5. Cela est moins vrai pour les systèmes utilisant la VBAP et ayant un grand nombre de haut-parleurs.

La WFS, quand à elle, impose de répartir les haut-parleurs à intervalles réguliers sur une ligne. Plusieurs exemples d'intégration de systèmes WFS avec des dispositifs de rendu visuel sont présentés dans la monographie de de Vries [2009]. Les interfaces visuelles imposent donc certains aménagements des technologies de *rendu sonore spatialisé* qui réduisent inévitablement la qualité du rendu sonore final.

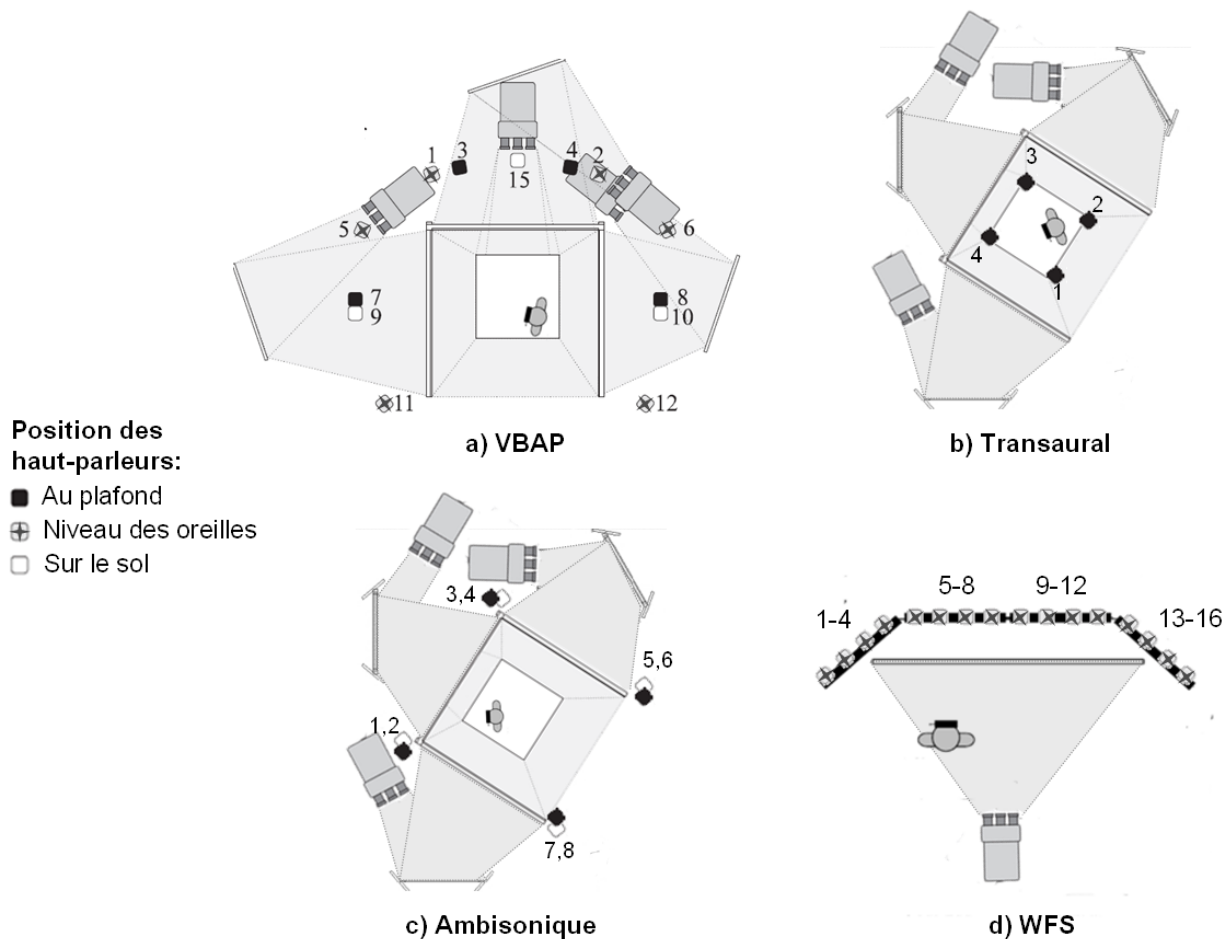


FIG. 2.5 – Exemples d'intégration de systèmes *externes* avec des interfaces visuelles de type CAVE : Systèmes réalisant un *rendu sonore spatialisé* par a) VBAP [Hiipakka et al. 2000, 2001], b) Transaural [Kuhlen et al. 2007], c) Ambisonique [Faria et al. 2005a], d) WFS [Springer et al. 2006].

En second lieu, comme le montrent Hiipakka et al. [2000, 2001], Ogi et al. [2003], Okubo et al. [2005], Conant et al. [2008], les écrans de projection vont avoir une influence non-négligeable sur le *rendu sonore spatialisé* créé par les technologies *holophoniques*. Des écrans rigides formant tout ou une partie d'un cube vont inévitablement créer des réflexions acoustiques qui vont nuire à la synthèse physique du champ acoustique que cherche à réaliser l'*holophonie*. Des écrans souples permettent de pallier partiellement ce défaut, mais induisent quand même des déformations spatio-temporelles tout aussi nuisibles du champ acoustique généré par les haut-parleurs. En particulier, une atténuation considérable des fréquences supérieures à 2 kHz a été mesurée par Hiipakka et al. [2000, 2001] pour des écrans souples. Ces défauts peuvent partiellement être compensés par une égalisation dans le domaine temporel ou fréquentiel [Corteel 2006], mais le *rendu sonore spatialisé* reste aussi inévitablement dégradé par la présence des écrans.

En résumé, les technologies *holophoniques* semblent être les plus intéressantes, car elles constituent le meilleur compromis entre la taille de la zone d'écoute et la qualité de la restitution spatiale. Cependant, l'intégration des technologies *holophoniques* dans des systèmes de restitution visuelle basés sur des écrans de grandes dimensions se heurte à deux problèmes pratiques : les conflits de positionnement entre les écrans et les haut-parleurs, et le filtrage du champ sonore induit par les écrans de projection. La section suivante présente alors un dispositif permettant d'intégrer un système de restitution spatialisée du son de type *holophonique* avec un système de rendu visuel basé sur des écrans de grandes dimensions tout en s'affranchissant de ces contraintes pratiques.

2.5 Le SMART-I²

Dans cette section un dispositif de rendu audio-visuel spatialisé pour la réalité virtuelle, le SMART-I² («*Spatial Multi-user Audio-visual Real-Time Interactive Interface*»), est présenté. Ce dispositif a été développé au LIMSI au cours de la thèse et a fait l'objet d'une présentation lors de la 125^{ième} convention de l'AES (Audio Engineering Society) [Rébillat et al. 2008] ainsi que de présentations invitées lors des conférences *IEEE 3D-TV* [Rébillat et al. 2009] et *Euro-VR Eve 2010* [Rébillat et al. 2010a].

Une vue d'ensemble du système est d'abord proposée, puis les différents éléments technologiques qui font la spécificité du dispositif sont présentés : les «*Large Multi-Actuator Panels (LaMAPs)*», la «*Wave Field Synthesis (WFS)*» et la stéréoscopie passive avec suivi.

2.5.1 Spécificités du système

Le SMART-I² est un dispositif de rendu audio-visuel pour la réalité virtuelle qui a été pensé, en accord avec la définition de la réalité virtuelle donnée à la section 2.2.1, comme une *fenêtre audio-visuelle ouverte sur un monde virtuel*. Lors du développement du système, différents efforts ont été faits pour pallier les limitations des systèmes de réalité virtuelle évoquées précédemment (voir section 2.4.3). Pour maximiser l'*immersion* des utilisateurs dans le monde virtuel, une attention spéciale a été portée aux deux points suivants :

- **Cohérence audio-visuelle spatiale** : le *rendu sonore spatialisé* a été réalisé en utilisant la «*Wave Field Synthesis*» (WFS, voir section 2.5.3) couplée à une simulation de réverbération. Cela permet la synthèse physique d'un champ sonore dans une large zone de l'espace et réalise donc un *rendu sonore spatialisé* complet (direction, distance, effet de parallaxe et réverbération) dans une large zone de restitution. La mise en relation du *rendu sonore spatialisé* réalisé par la WFS avec le rendu visuel tridimensionnel permet ainsi de créer pour l'utilisateur un rendu audio-visuel spatialement cohérent dans une large zone de l'espace. Cette spécificité du SMART-I² est susceptible d'avoir un impact favorable sur différents aspects de la perception d'un environnement audio-visuel de réalité virtuelle (voir tableau 2.1).
- **Intégration dans l'interface visuelle** : Pour réaliser le SMART-I², une technologie spécifique de haut-parleurs, les «*Large Multi-Actuator Panels*» (LaMAPs, voir section 2.5.4), a été utilisée. Elle permet d'intégrer à la fois les écrans et les haut-parleurs au sein d'un même élément technologique,

en s'affranchissant ainsi d'une part des conflits de positionnement entre les haut-parleurs, les vidéo-projecteurs et les écrans, d'autre part des détériorations spatio-temporelles du champ sonore causées par les écrans.

2.5.2 Vue d'ensemble

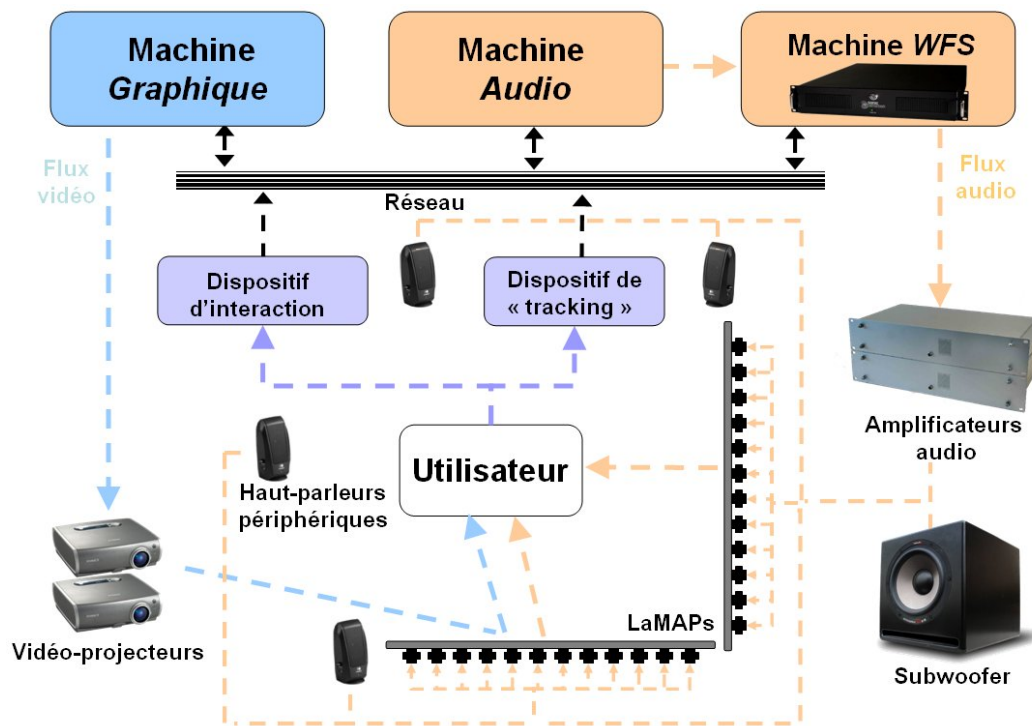


FIG. 2.6 – Schéma général de l'organisation physique et informatique du SMART-I².

Le système est physiquement composé de deux LaMAPs (2.6 m × 2 m) formant un coin de CAVE. Cela permet l'évolution des utilisateurs dans une large zone de rendu (2.5 m × 2.5 m) restituant un grand champ de vision (150° en horizontal et 70° en vertical) et d'audition (150° en horizontal) du monde virtuel. Chaque LaMAP est équipé de 12 excitateurs électro-dynamiques espacés de 21 cm. Pour pallier le manque de réponse des LaMAPs en basses fréquences un *subwoofer* a été adjoint au système. De plus 6 haut-parleurs électro-dynamiques sont placés autour de la zone de restitution pour la simulation de la réverbération. La position des utilisateurs dans la zone de restitution est suivie en temps réel grâce à un dispositif de *tracking* optique. Les utilisateurs ont éventuellement accès à divers moyens d'interaction (Wiimote, etc.). Une vue d'ensemble du système est présentée sur la figure 2.6.

Pour le rendu audio-visuel en temps réel, une approche distribuée est utilisée. Les *flux audio* correspondant à chacun des objets audio virtuels sont générés par la machine *audio* et envoyés à la machine *WFS* mise à notre disposition par *sonic emotion*. Cette dernière calcule alors avec une très faible latence les signaux à envoyer à chacun des 24 excitateurs constitutifs des LaMAPs pour créer un *rendu sonore spatialisé* par WFS. Les signaux acoustiques correspondant à la *réverbération* générée par chacun des objets virtuels

présents dans la scène audio-visuelle sont générés sur la machine *audio* par le *Spatialisateur*⁶ de l'Ircam [Jot 1999] et restitués directement par les haut-parleurs périphériques. En parallèle, la machine *graphique* génère le *flux vidéo* correspondant. Toutes les composantes du rendu audio-visuel communiquent entre elles par un réseau au sein duquel transite la description de la scène audio-visuelle (position des objets audio-visuels, position de l'utilisateur, paramètres de l'interaction, marche/arrêt, etc ...). Ces informations sont utilisées pour mettre à jour en temps réel les rendus audio et visuel, ce qui rend le système pleinement interactif.

2.5.3 Rendu sonore spatialisé par «Wave Field Synthesis»

La «Wave Field Synthesis» (WFS) est la technologie de rendu spatialisé du son qui fait la spécificité du SMART-I². Ses fondements théoriques ainsi que sa mise en œuvre sont ici présentés d'après Berkhout et al. [1993], Corteel [2004, 2006], de Vries [2009]. Cette technologie crée un *rendu sonore spatialisé* tel que décrit à la section 2.3.1 capable de restituer *direction*, *distance* et *effet de parallaxe* dans une large zone de rendu. La *réverbération* est ensuite générée grâce au *Spatialisateur* de l'Ircam et aux haut-parleurs périphériques.

Théorie de la WFS

La «Wave Field Synthesis» s'appuie sur le principe de Huygens-Fresnel pour restituer un champ de pression donné dans une zone étendue de l'espace. Ce principe peut être formulé comme suit [Bruneau & Scelo 2006, p. 315] :

«Chaque point M recevant une onde issue d'une **source primaire** se comporte comme une **source secondaire** rayonnant une onde sphérique de même fréquence, même amplitude et même phase que l'onde reçue. Les fronts d'onde créés par les **sources secondaires** interfèrent entre eux pour recréer un front d'onde identique à celui émis par la **source primaire**.»

Soient un volume de reproduction Ω_R et un volume Ω_Ψ où est située une source primaire P , tous deux séparés par une surface fermée $\partial\Omega$. Il est supposé que la source primaire P engendre en M le champ de pression $p_\Psi(M, t)$. Le principe de Huygens-Fresnel énonce qu'il est possible de reproduire dans le volume Ω_R le champ de pression $p_\Psi(M, t)$ issu de la source P contenue dans Ω_Ψ , à condition que l'on connaisse à tout instant l'amplitude et la phase de $p_\Psi(M, t)$ en tout point de la surface $\partial\Omega$ séparant les deux volumes (voir figure 2.7 pour les notations).

Le principe de Huygens-Fresnel consiste donc en une formulation intégrale, dans le domaine fréquentiel, donnant le champ de pression $P_R(M, \omega)$ engendré en un point M de l'espace de restitution Ω_R , par les sources secondaires réparties sur $\partial\Omega$ [Berkhout et al. 1993] :

$$P_R(M, \omega) = \iint_{\partial\Omega} \left(P_\Psi(S, \omega) \vec{\nabla}[G(M, S)] - G(M, S) \vec{\nabla}[P_\Psi(S, \omega)] \right) \cdot d\vec{\Omega} \quad (2.1)$$

avec :

- $P_R(M, \omega)$: la transformée de Fourier de la pression $p_R(M, t)$ engendrée par les sources secondaires au point d'écoute M .

6. <http://forumnet.ircam.fr/692.html>

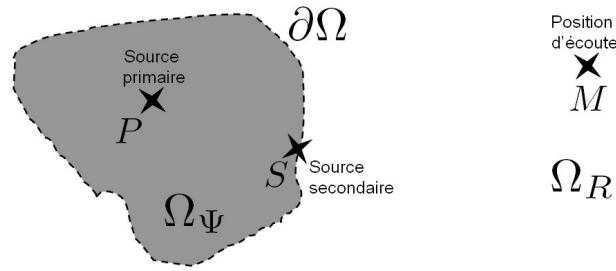


FIG. 2.7 – Définition de l'espace des sources primaires Ω_Ψ , de la surface de sources secondaires $\partial\Omega$, et de l'espace de reproduction Ω_R .

- $P_\Psi(S, \omega)$: la transformée de Fourier de la pression $P_\Psi(S, \omega)$ engendrée par la source primaire P en un point de la surface $\partial\Omega$ correspondant à une source secondaire S .
- $G(M, S) = \frac{\exp(-jk||\vec{M}\vec{S}||)}{4\pi||\vec{M}\vec{S}||}$: la fonction de Green associée aux problèmes tridimensionnels de propagation sonore en espace infini et satisfaisant aux conditions de Sommerfeld à l'infini, dans le domaine fréquentiel.

L'intégrale de l'équation (2.1) vaut par définition :

$$P_R(M, \omega) = \begin{cases} P_\Psi(M, \omega) & \text{pour tous les points de } \Omega_R \\ 0 & \text{pour tous les points de } \Omega_\Psi \end{cases} \quad (2.2)$$

ce qui assure donc la reconstruction du champ sonore engendré par la source primaire P en tout point M de l'espace de reproduction Ω_R .

L'équation (2.1) montre donc que le champ rayonné dans l'espace de restitution peut être considéré comme la superposition :

- d'un champ rayonné par une distribution surfacique de sources secondaires de type *monopôles* alimentés par la composante normale à $\partial\Omega$ du gradient du champ de pression de la source.
- d'un champ rayonné par une distribution surfacique de sources secondaires de type *dipôles* alimentés par l'amplitude du champ de pression de la source.

De plus, dans le cas particulier où la surface $\partial\Omega$ est un plan infini, il est possible de réduire l'intégrale à l'un ou l'autre des types de sources [Corteel 2004]. Les éléments électro-acoustiques actuels étant plutôt des monopôles, c'est la représentation monopolaire qui est choisie pour les sources secondaires. Alors l'équation (2.4), aussi appelée «*Équation de Rayleigh I*», est obtenue :

$$P_R(M, \omega) = -2 \iint_{\partial\Omega} G(M, S) \vec{\nabla}[P_\Psi(S, \omega)] \cdot d\vec{S} \quad (2.4)$$

Cette équation décrit le champ de pression dans l'espace de restitution comme étant issu d'une répartition surfacique continue de sources secondaires monopolaires disposées sur un plan infini.

À partir de cette formulation du principe de Huygens-Fresnel, il est alors théoriquement possible de reproduire dans une zone étendue de l'espace un champ sonore créé par une ou plusieurs sources primaires à partir d'une répartition surfacique continue de sources secondaires monopolaires.

Réduction à une ligne infinie de sources secondaires

En pratique, il n'existe pas de répartition surfacique plane, infinie et continue de sources secondaires monopôlaires contrôlable en tout point. On est donc amené à réaliser des approximations pour mettre en œuvre le principe de restitution sonore décrit précédemment [Corteel 2006].

En premier lieu, la répartition surfacique de sources secondaires est réduite à une distribution linéique. Cette réduction n'est pas obligatoire mais est motivée par les arguments suivants [Corteel 2006] :

- La tête de l'auditeur est dans la majorité des cas dans le plan des sources qu'il écoute.
- La localisation dans le plan horizontal est perceptivement la plus pertinente (voir section 7.2).
- La complexité du système est réduite si le nombre de sources secondaires est réduit.

Cette approximation a en pratique trois conséquences directes :

- Le *rendu sonore spatialisé* est **limité au plan horizontal**. Avec cette formulation, il est donc impossible de simuler l'élévation des sources sonores.
- Le champ de pression rayonné est à **symétrie cylindrique**, autour de l'axe que constitue la ligne sur laquelle sont réparties les sources secondaires.
- **L'atténuation du champ sonore n'est pas correcte dans la zone de rendu** et n'est bien restituée que sur une ligne parallèle à la ligne que constituent les sources secondaires.

Réduction à une ligne finie de sources secondaires

Pour des raisons évidentes d'encombrement, il n'est pas non plus possible d'avoir une ligne infinie de sources secondaires. La zone où se situe les sources secondaires est en pratique réduite à un segment, ce qui pose deux problèmes pratiques :

- **Diffraction** : Le segment de sources sonores constitue une fenêtre acoustique ouverte entre deux espaces sonores. Étant donnée la taille des longueurs d'onde associées aux phénomènes acoustiques (de quelques centimètres à plusieurs mètres !), il y a donc apparition d'un phénomène de diffraction par cette fenêtre qui parasitera le champ acoustique que l'on cherche à obtenir.
- **Limitation de zone d'écoute** : La limitation de la distribution de sources à un segment va limiter la taille de la zone d'écoute dans laquelle il est possible de restituer correctement les sources ainsi que la taille de l'espace virtuel dans lequel il va être possible de placer les sources [Corteel 2006]. Sur la figure 2.8 la ligne de sources secondaire définit une *fenêtre sonore* ouverte sur le monde virtuel. Par conséquent, en parcourant la zone de restitution Ω_R donnée, seule une partie de l'espace virtuel sonore est *visible* (ou plus exactement *audible*) par l'utilisateur. La zone d'existence des sources primaires Ω_Ψ désigne alors l'ensemble des sources primaires qui sont *visibles* depuis tous les points de la zone de restitution Ω_R à travers la ligne de sources secondaires.

Échantillonnage spatial de la ligne finie de sources secondaires

La dernière approximation réalisée en pratique est un échantillonnage spatial de la distribution linéique finie de sources secondaires, comme schématisé sur la figure 2.8. Une longueur d'onde spatiale de repliement λ_a en deçà de laquelle le champ sonore n'est pas correctement reconstruit existe donc pratiquement. Étant donné qu'*espace* et *temps* sont couplés par l'intermédiaire de l'équation de propagation des ondes

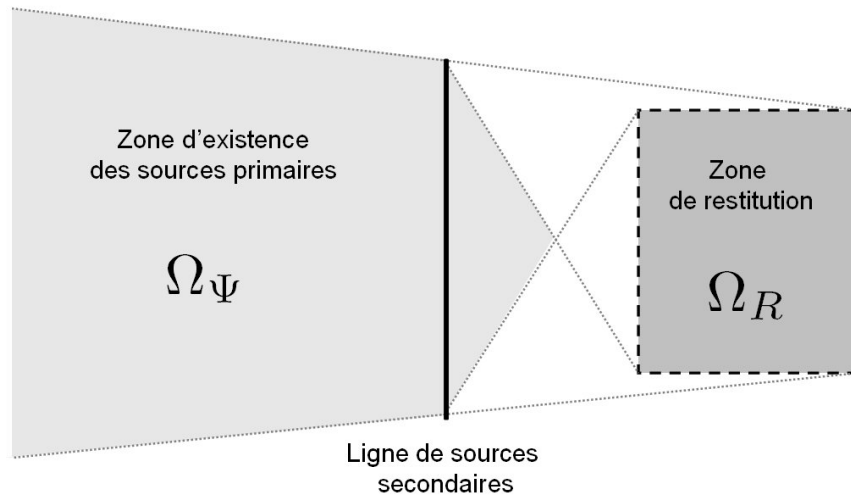


FIG. 2.8 – Positions possibles des sources primaires en fonction des dispositions de la ligne de sources secondaires et de la zone d'écoute.

sonores dans l'air, une fréquence temporelle f_{al} liée au repliement spatial existe donc aussi. Cette fréquence dépend de la position de l'auditeur, de la position de la source primaire, et augmente avec la distance entre l'auditeur et les sources secondaires. Cette fréquence est donnée par Corteel [2006], Corteel et al. [2008] :

$$f_{al}(P, M) = \frac{1}{\max_{S \in [1..N]} \Delta t_S(P, M)} \quad (2.5)$$

où :

- P désigne la position de la source primaire à laquelle on se réfère (voir figure 2.7).
- M désigne la position d'écoute (voir figure 2.7).
- S désigne une source secondaire parmi N .
- $\Delta t_S(P, M) = t_{S+1}(P, M) - t_S(P, M)$ représente la différence des durées de propagation de sons émis simultanément par deux sources secondaires adjacentes.

Les effets de l'échantillonnage spatial de la ligne finie de sources secondaires sont illustrés sur la figure 2.9. Sur la figure de gauche, les réponses impulsionnelles temporelles d'un banc de 60 sources secondaires et long de 6 m synthétisant une source primaire centrée située 6 m derrière le banc sont mesurées à deux positions d'écoute [Corteel 2006]. Les réponses impulsionnelles font toutes deux apparaître d'abord un premier front d'onde auquel toutes les sources secondaires contribuent en phase et qui correspond à celui que l'on désire synthétiser. Ensuite apparaissent les contributions individuelles de chacune des sources secondaires. Sur la figure de droite, les réponses impulsionnelles d'un banc composé de neuf sources secondaires et long de 2 m reproduisant une onde plane se propageant perpendiculairement au banc de sources secondaires ont été mesurées, sur une ligne de positions d'écoute parallèle au banc de haut-parleurs (d'après [Boone 2004]). Les réponses impulsionnelles temporelles (axe vertical) sont tracées pour les différentes positions d'écoute (axe horizontal). L'amplitude de la réponse impulsionnelle à un instant donné et à une position donnée est codée par le niveau de gris. Dans cette représentation spatio-temporelle, des fronts d'ondes dus aux contributions individuelles des sources secondaires se propagent en arrière du front d'onde principal correspondant à la source primaire. Le phénomène de repliement spectral du à l'espace est de fa-

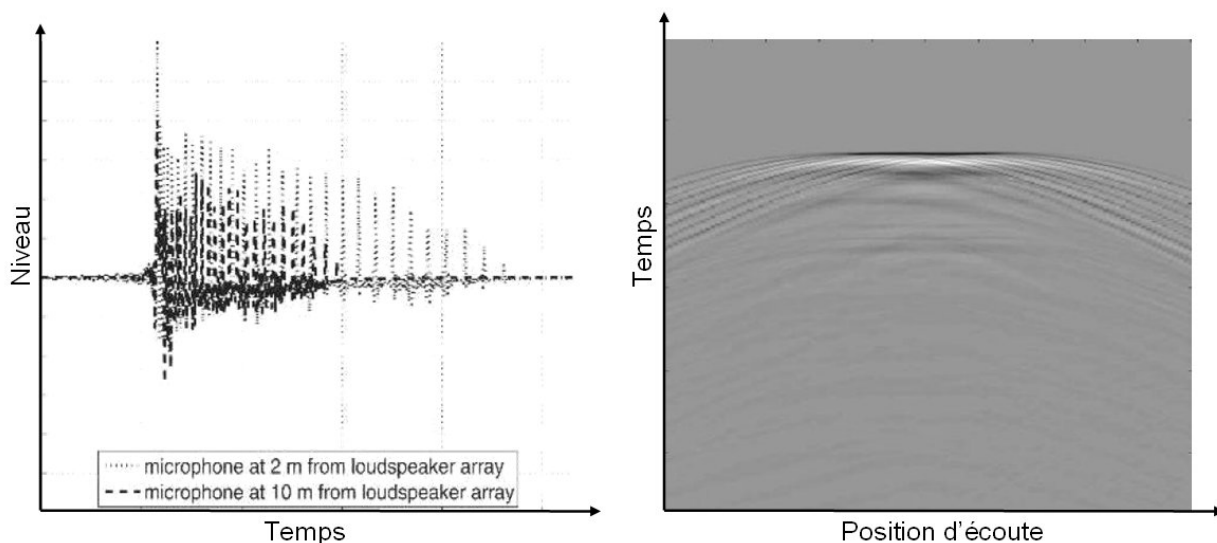


FIG. 2.9 – Illustration des effets de l'échantillonnage spatial.

À gauche : réponses impulsionnelles temporelles d'un banc de 60 sources secondaires et long de 6 m synthétisant une source primaire centrée située 6 m derrière lui. Les réponses impulsionnelles mesurées à deux positions d'écoute différentes sont tracées (extrait de [Corteel 2006]). Les axes sont indiqués à titre indicatif et par conséquent dénués de graduations.

À droite : les réponses impulsionnelles d'un banc composé de 9 sources secondaires et long de 2 m reproduisant une onde plane se propageant perpendiculairement au banc de sources secondaires ont été mesurées sur une ligne de positions d'écoute parallèle au banc de haut-parleurs. Les réponses impulsionnelles temporelles sont tracées pour les différentes positions d'écoute. L'amplitude de la réponse impulsionnelle à un instant donné et à une position donnée est codée par le niveau de gris. (extrait de [Boone 2004]).

çon assez générale complexe à décrire à cause de sa dépendance spatio-temporelle [Pueo et al. 2007].

Pratiquement, un écart de 15 à 20 cm entre deux sources secondaires consécutives conduit à une fréquence temporelle liée au repliement spatial $f_{al} \approx 1500$ Hz. Les conséquences perceptives de cette fréquence de repliement spectral sont discutées à la section 8.2.4.

Égalisation des sources secondaires

Le dernier point à prendre en compte est la façon dont rayonnent les sources secondaires réelles dans l'espace en fonction de la fréquence. En effet, la théorie développée précédemment fait l'hypothèse que le champ sonore est reproduit à l'aide d'une distribution de sources secondaires *monopôlaires*. Les sources secondaires sont donc supposées parfaitement omnidirectionnelles et à réponse fréquentielle plate. Or, elles ne présentent en pratique jamais de caractéristiques aussi idéales. Pour essayer de minimiser l'influence de ces défauts dans le contexte de la WFS, des techniques dites d'*égalisation* ont été employées.

La technique d'égalisation classique consiste à considérer séparément chacune des sources secondaires et à essayer de compenser les défauts de sa réponse fréquentielle moyennée en différents points de l'espace. Avec cette technique, on ne prend cependant que très peu en compte les caractéristiques spatiales de rayonnement des émetteurs ni le fait que l'objectif soit la reconstruction d'un champ de pression acoustique émanant d'une source sonore virtuelle.

Une autre technique d'égalisation utilise le *filtrage inverse multicanal* [Corteel 2006]. Cette technique

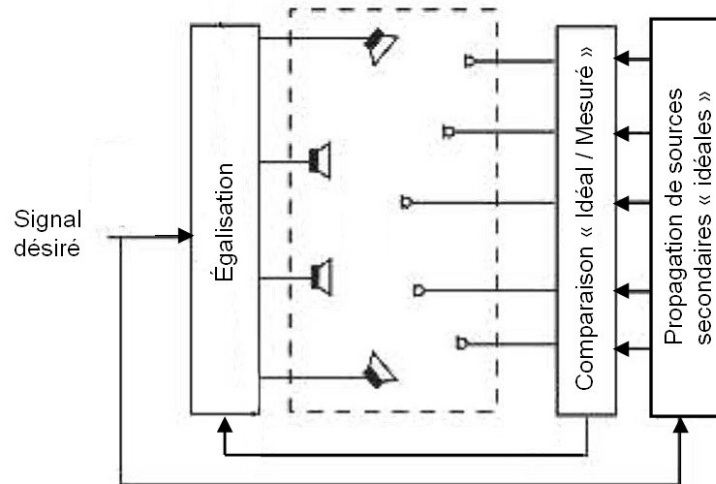


FIG. 2.10 – Schéma illustrant le principe du *filtrage inverse multicanal* (d'après [Corteel 2006]).

modélise le problème comme un système à plusieurs entrées (les sources secondaires) et plusieurs sorties (des microphones disposés à des points où l'on souhaite contrôler le champ de pression). Connaissant les signaux d'entrée, la mesure des signaux obtenus en chacune des sorties donne accès à une matrice de transfert *entrée/sortie*. D'autre part, connaissant la réponse idéale désirée en chacun des points de sortie il est alors possible de calculer l'erreur entre le modèle et les mesures et d'estimer ainsi les filtres à appliquer pour minimiser cette erreur (cf. figure 2.10). Cette technique permet ainsi de contrôler le champ de pression acoustique en un nombre fini de points de la zone de rendu. Le *filtrage inverse multicanal* prend donc en compte d'une part les caractéristiques spatiales de rayonnement des émetteurs et d'autre part l'objectif de reconstruire un champ sonore, mais d'un seul bloc et de façon aveugle. Ce type d'égalisation ne garantit donc en aucun cas que le champ sonore sera correctement reconstruit en dehors des points de contrôle.

Des techniques de *filtrage inverse multicanal* avancées ont donc été développées pour pallier les problèmes évoqués au paragraphe précédent [Corteel 2006]. La procédure utilisée par ces techniques est similaire à celle décrite précédemment, sauf qu'elle est à présent combinée à une description physique du champ sonore émis par le banc de haut-parleurs. Les caractéristiques spatiales de rayonnement des émetteurs et l'objectif de reconstruction d'un champ sonore sont ainsi pris en compte de façon séparée. Ces méthodes *informées* permettent donc de garantir une reconstruction correcte du champ de pression acoustique dans une zone de l'espace plus large que les seuls points de contrôle.

2.5.4 Les *Large Multi-Actuator Panels* (LaMAPs)

Dans les systèmes dédiés à la réalité virtuelle, les contraintes imposées sur le placement des haut-parleurs et les déformations du champ acoustique induites par les écrans nuisent généralement à la qualité du *rendu sonore spatialisé* obtenu (voir section 2.4.3). D'un point de vue technologique, il est donc intéressant de chercher à fabriquer des «haut-parleurs plans qui puissent être utilisés comme des écrans» [Bai & Chung 2002, Pueo et al. 2008, Lopez et al. 2008].

Le concept *haut-parleurs plans* existe depuis relativement longtemps (voir Pueo et al. [2010a] pour une

synthèse bibliographique sur le sujet). Un haut-parleur plan est typiquement constitué d'un panneau léger et rigide au dos duquel est fixé un excitateur électrodynamique (voir figure 2.11). L'excitateur électrodynamique génère des ondes mécaniques de flexion dans le panneau, et ces ondes ont un rayonnement acoustique dans l'espace environnant. Initialement, cette technologie nommée «*Distributed Mode Loudspeakers*» (DML) fait intervenir un ou plusieurs excitateurs électro-dynamiques placés de façon à exciter le plus grand nombre possible de modes. Tous les actuateurs sont pilotés par le même signal audio [Bank 1998, Angus 1999] (voir figure 2.11(a)). L'idée selon laquelle les différents excitateurs pouvaient être pilotés indépendamment a ensuite émergé pour donner naissance aux «*Multi-Actuator Panels*» (MAPs) [Corteel et al. 2002, Horbach et al. 2002] (voir figure 2.11(b)). Cependant, pour que les MAPs puissent être intégrés efficacement avec une interface visuelle destinée à la réalité virtuelle, une grande surface de projection est nécessaire. La surface du panneau constituant ces MAPs a donc été considérablement agrandie pour donner lieu à ce que nous appelons ici des «*Large Multi-Actuator Panels*» (LaMAPs) (voir figure 2.11(c)).

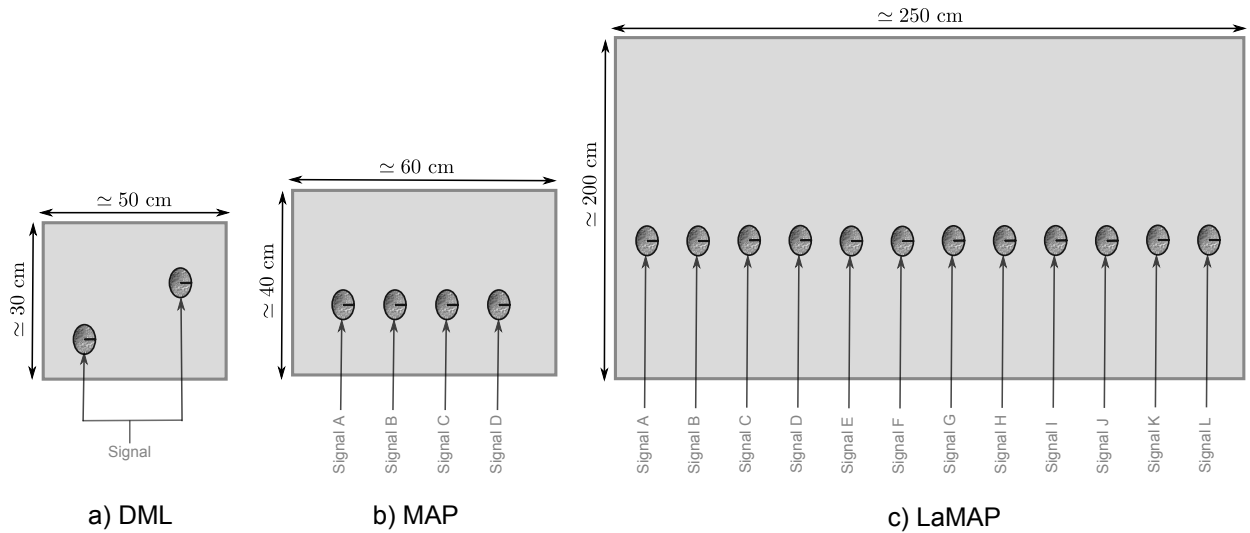


FIG. 2.11 – Du DML au LaMAP : a) *Distributed Mode Loudspeaker* (DML). b) *Multi-Actuator Panel* (MAP) c) *Large Multi-Actuator Panel* (LaMAP).

Les MAPs et LaMAPs possèdent aussi un avantage pratique majeur par rapport aux haut-parleurs électrodynamiques classiques en ce qui concerne leur application à la WFS. En effet, il est plus facile d'obtenir une ligne de sources secondaires dense en utilisant des MAPs ou des LaMAPs qu'en utilisant des haut-parleurs électrodynamiques, qui sont plus volumineux. Les technologies des MAPs et des LaMAPs ont donc récemment été utilisées avec succès pour générer des *rendus sonores spatialisés* basés sur la WFS [Boone 2004, Corteel et al. 2007, Lopez et al. 2008, Pueo et al. 2008, 2009, 2010b,a].

Le désavantage des MAPs, par rapport à une interface visuelle pour la réalité virtuelle utilisant de la rétro-projection, est qu'ils ne sont pas visuellement transparents. L'utilisation de LaMAPs impose alors une projection frontale des images, ce qui pose des problèmes d'ombre portée par le participant sur l'écran et donc limite pratiquement la taille de la zone de rendu visuel.

Ainsi, le SMART-I² utilise les LaMAPs pilotés par la WFS pour la génération du *rendu sonore spatialisé*. La face avant de ces mêmes LaMAPs sert quand à elle d'écran de projection. Des vues des faces avant et arrière du SMART-I² sont présentées sur la figure 2.12. À ce jour, seul un autre prototype de LaMAP piloté

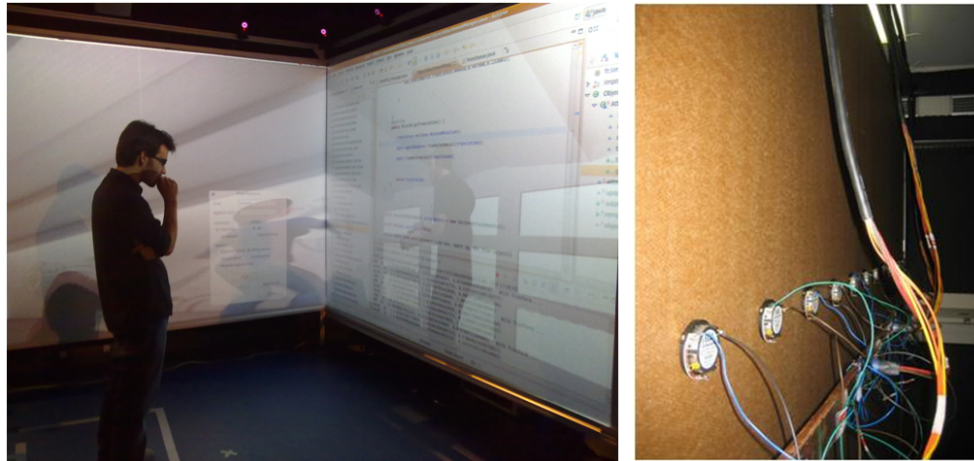


FIG. 2.12 – Faces avant et arrière des LaMAPs composant le SMART-I².

par WFS a été développé et testé par des chercheurs de l'université d'Alicante en Espagne [Lopez et al. 2008, Pueo et al. 2008]. Ce prototype n'a semble-t-il pas été combiné avec un système de rendu visuel. Le SMART-I² serait alors le premier dispositif audio-visuel pour la réalité virtuelle conçu à partir de LaMAPs qui soit effectivement opérationnel.

2.5.5 Stéréoscopie passive avec suivi

Pour créer un rendu visuel tridimensionnel convaincant et cohérent avec le *rendu sonore spatialisé*, il faut d'une part être en mesure de générer et de projeter une scène visuelle spatialement interprétable par les utilisateurs, et aussi être capable de créer sur chaque œil le stimulus visuel qui lui est destiné. Cette section est largement inspirée des écrits de Fuchs [2006].

Principe de la stéréoscopie passive

La stéréoscopie passive est une technologie utilisant les propriétés de polarisation de la lumière pour créer sur chaque œil de l'utilisateur le stimulus qui lui est destiné. Cette technologie consiste à projeter simultanément sur un écran métallisé⁷ les deux images destinées aux yeux de l'utilisateur et polarisées orthogonalement l'une par rapport à l'autre. L'utilisateur regarde ensuite cet écran avec des lunettes présentant des filtres polarisants similaires à ceux utilisés pour la projection. Des images différentes peuvent alors être envoyées à chacun des yeux de l'utilisateur (voir figure 2.13). Cependant, les filtres polarisants n'étant pas parfaits, la réjection de l'image non-désirée d'un œil sur l'autre n'est pas totale et des images dites *fantômes* peuvent alors apparaître. Dans le SMART-I², des filtres polarisés linéairement sont utilisés à cet effet.

Création du rendu visuel spatialisé

La seconde étape dans la création d'un rendu visuel tridimensionnel consiste à créer les deux images à envoyer à chaque instant aux deux yeux. Au quotidien, les yeux perçoivent deux images spatialement

7. La fine pellicule métallique qui recouvre les écrans permet de conserver la polarisation de la lumière.

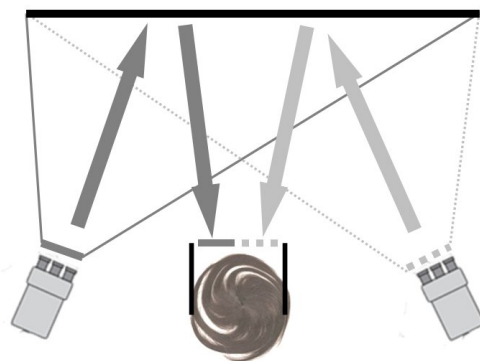


FIG. 2.13 – Principe de la projection polarisée.

Les projecteurs gauche et droite sont couplés à des filtres polarisés orthogonaux l'un avec l'autre. L'utilisateur porte des lunettes où chacun des yeux est équipé de ces mêmes filtres. Ainsi, l'image projetée par le vidéoprojecteur de gauche n'est vue que par l'œil gauche et celle projetée par le vidéoprojecteur de droite que par l'œil droit.

légèrement décalées d'une même scène visuelle. Le cerveau humain interprète ensuite ces deux images pour se forger sa propre représentation visuelle tridimensionnelle du monde environnant dans lequel il peut évoluer. Ce sont ces deux images qu'il faut alors créer de telle sorte que l'utilisateur perçoive le monde virtuel aux dimensions désirées.

Il est judicieux de commencer par s'intéresser à un seul œil dans le cas de la réalité virtuelle (voir figure 2.14). Le monde virtuel est alors observé depuis un point de vue donné (caméra virtuelle située à la position de l'œil) et l'image correspondant à ce point de vue est projetée sur un plan virtuel associé à la caméra virtuelle. L'image plane captée par la caméra virtuelle est ensuite envoyée à l'un des vidéoprojecteurs qui l'affiche sur l'écran de projection, qui lui, est bien réel. L'utilisateur voit alors l'image projetée et se forge ainsi sa propre représentation visuelle spatiale du monde virtuel. Pour l'autre œil, le procédé est rigoureusement le même, seul le point de vue dans le monde virtuel est légèrement différent.

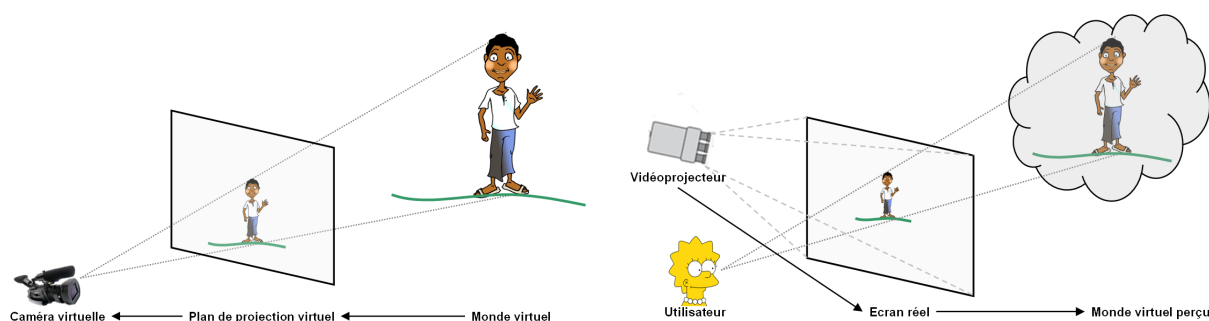


FIG. 2.14 – Principe de la restitution visuelle du monde virtuel

Ce procédé de calcul des deux images à envoyer à chaque œil couplé avec la technologie de stéréoscopie passive permet donc de créer des objets visuels virtuels perçus dans l'espace virtuel par l'utilisateur devant ou derrière l'écran.

Limites de l'utilisation d'un écran de projection

L'utilisation de cette technologie de rendu visuel ne permet cependant pas une restitution fidèle de la réalité [Wann et al. 1995, Hoffman et al. 2008, Howarth 2011]. En effet, lorsque qu'une personne regarde un objet dans le monde réel, ses yeux cherchent à voir l'objet net (*accommodation*) et leur axes optiques convergent vers l'objet (*vergence*, voir aussi la section 7.2.2). Dans le monde réel, l'*accommodation* et la *vergence* sont donc couplées et dépendent de la distance de l'objet qui est vu, voir figure 2.15(a). Dans le cas d'une personne regardant un objet virtuel via le SMART-I², ses yeux cherchent à voir l'écran net (et non pas l'objet virtuel) mais leurs axes optiques convergent vers l'objet virtuel. Dans le monde virtuel l'*accommodation* et la *vergence* sont donc découplées, voir figure 2.15(b).

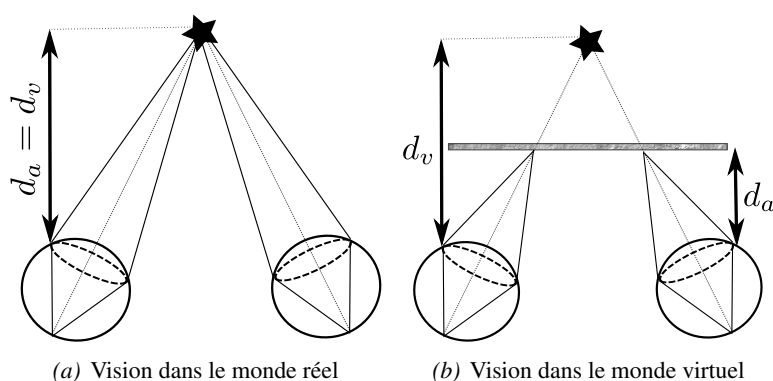


FIG. 2.15 – Illustration des limitations de la projection stéréoscopique sur écran.

S'il reste *raisonnable*, ce découplage ne nuit cependant pas à la vision binoculaire. Plus précisément si l'écart entre les distances de vergence (d_v) et les distances d'accommodation (d_a) n'excède pas ± 0.3 D⁸ alors la vision binoculaire est confortable (zone de Percival) [Hoffman et al. 2008]. De plus, pour que l'utilisateur fusionne en un seul objet 3D les deux images qui lui sont présentées, il est nécessaire que l'écart angulaire existant entre la direction vers laquelle est réalisé l'*accommodation* et celle où convergent les axes optiques n'excède pas $\frac{15}{60}^\circ$ (zone de Panum) [Hoffman et al. 2008].

Sur la figure 2.16, la zone où la vision 3D est confortable (zone de Percival) et celle où les images binoculaires sont correctement intégrées (zone de Panum) sont tracées pour un objet réel et pour un objet virtuel restitué via le SMART-I² à un utilisateur situé à 2 m du dispositif. Globalement, la zone de confort est plus étroite que la zone d'intégration binoculaire. Cela indique qu'il est possible d'intégrer des images présentant une assez grande disparité horizontale au prix d'un désagrément ou d'une fatigue visuelle perceptibles par les utilisateurs. De plus la figure 2.16 informe que pour un utilisateur situé à 2 m du SMART-I², il est possible de restituer confortablement des objets visuels situés entre 1.5 m et 2.8 m et que seuls les objets situés entre 1.3 m et 5 m seront correctement perçus en 3D. Par ailleurs, plus l'utilisateur est prêt du dispositif, plus l'étendue de ces zones diminue.

8. Dioptries (en m^{-1})

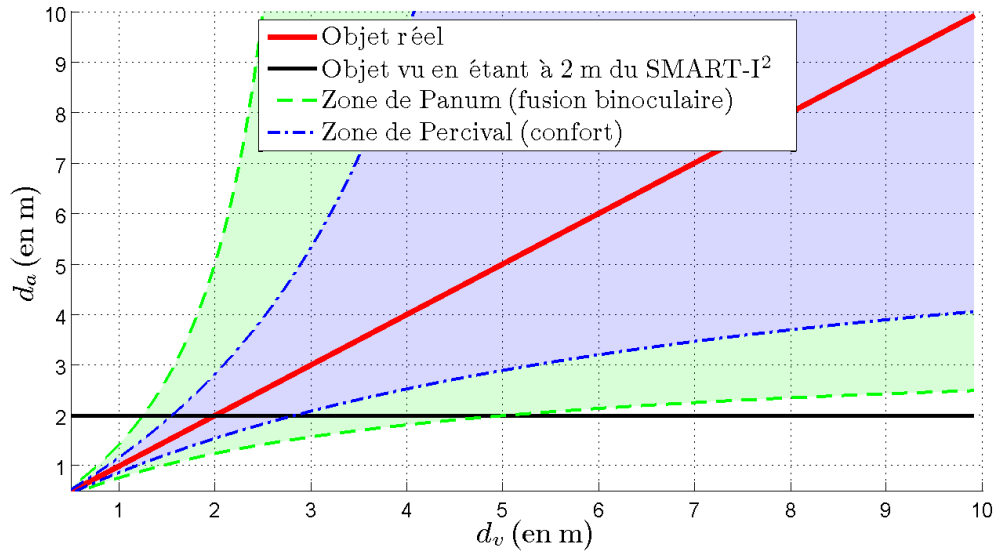


FIG. 2.16 – Zones où la vision 3D est confortable (zone de Percival) et où les images binoculaires sont correctement intégrées (zone de Panum) pour un objet réel et pour un objet virtuel restitué via le SMART-I² à un utilisateur situé à 2 m du dispositif. La distance de vergence (distance à laquelle converge les axes optiques des deux yeux) est noté d_v . La distance d'accommodation (distance à laquelle l'image vue par les yeux est nette) est notée d_a . Ces zones sont tracées pour un utilisateur ayant un écart inter-pupille de 6 cm.

Nécessité d'un système de suivi de l'utilisateur

Lorsqu'un utilisateur évolue dans la zone de restitution ou lorsque deux utilisateurs visualisent le même rendu visuel généré par la méthode précédente, des mouvements non désirés des objets apparaissent (cf. figure 2.17). Ces mouvements sont appelés mouvements *pseudoscopiques* et sont dus au fait que la projection est plane et que le point de vue proposé aux utilisateurs est fixe. Dans une interface visuelle de type CAVE, l'objectif étant de proposer aux utilisateurs une large zone de restitution dans laquelle il puisse évoluer, ces mouvements parasites sont très gênants. Pour réduire cet inconvénient, il est alors nécessaire de suivre en temps réel la tête de l'utilisateur pour adapter à chaque instant le rendu visuel au point de vue de l'utilisateur dans le monde virtuel. Couplé à la stéréoscopie passive, le suivi de la tête de l'utilisateur permet ainsi à un utilisateur de bénéficier toujours du bon point de vue.

La difficulté est ensuite de s'assurer que l'espace décrit par le monde virtuel est perçu par l'utilisateur aux dimensions désirées par les concepteurs du monde virtuel. Cela est primordial pour s'assurer que l'espace virtuel visuel est bien spatialement cohérent avec l'espace virtuel sonore généré par le dispositif de *rendu sonore spatialisé*. Le seul paramètre de jeu disponible pour réaliser cet objectif est le plan de projection virtuel sur lequel l'image plane du monde virtuel est calculée pour un point de vue donné. Ainsi, la taille et la position de ce plan de projection doivent être ajustées en temps réel à partir de données géométriques comme la taille, la position de l'écran réel et la position de l'utilisateur pour que le monde virtuel visuel soit correctement perçu par l'utilisateur.

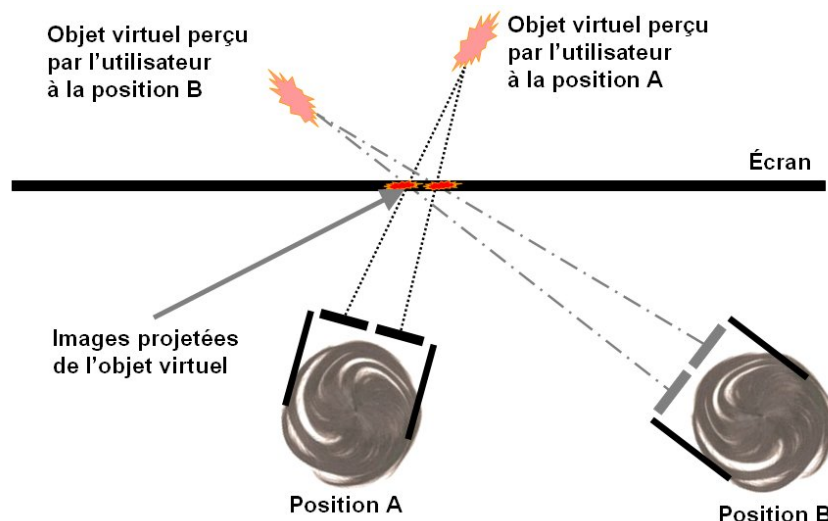


FIG. 2.17 – Illustration des mouvements *pseudoscopiques* : l'utilisateur perçoit l'objet visuel virtuel à deux positions différentes lorsqu'il se déplace de la position A à la position B.

2.6 Conclusion

L'objet du premier chapitre de cette thèse était de présenter le SMART-I² («*Spatial Multi-user Audio-visuel Real-Time Interactive Interface*»), un dispositif de restitution audio-visuelle spatialisée développé pour des applications de réalité virtuelle au cours de cette thèse. La notion de *réalité virtuelle* a d'abord été introduite en mettant l'accent sur la spécificité de l'approche à adopter lors de la conception d'un système destiné à la réalité virtuelle. La *présence*, qui est un facteur clé en réalité virtuelle, constitue l'une des quantités à optimiser. Une façon de maximiser la *présence* subjectivement perçue par les sujets est de proposer une interface qui soit la plus *transparente* et la plus *immersive* possible.

Dans l'état actuel de la science, un *rendu sonore spatialisé* est caractérisé par quatre attributs : la *direction*, la *distance*, la *réverbération* et l'*effet de parallaxe*. Les différents attributs d'un *rendu sonore spatialisé* ont un impact considérable, mais encore mal connu, sur un grand nombre de facteurs importants dans les applications de réalité virtuelle : présence, intelligibilité et qualité audio-visuelle globale, par exemple. La cohérence audio-visuelle spatiale semble de plus jouer un rôle important dans cet impact. Cela justifie de chercher à ajouter un *rendu sonore spatialisé* à un système *visuel* destiné à la réalité virtuelle.

Le SMART-I² propose alors une intégration originale d'un système de *rendu sonore spatialisé* de type *holophonique* (WFS) au sein d'un système de rendu *visuel* basé sur des écrans de grandes dimensions. Le schéma d'intégration proposé résout de plus deux problèmes pratiques : les conflits de positionnement entre les écrans et les haut-parleurs, et le filtrage du champ sonore induit par les écrans de projections. Le SMART-I² peut donc potentiellement permettre d'améliorer considérablement les performances de différentes applications de réalité virtuelle, mais aussi d'étudier les impacts des différents attributs du son spatialisé sur ces performances.

En guise de perspectives, il est possible de mentionner que le "M" du SMART-I², faisant référence à l'aspect multi-utilisateur, n'est en réalité que partiellement opérationnel. En effet, la stéréoscopie passive avec suivi, telle qu'elle est présentée dans la section 2.5.5, ne permet qu'à un seul utilisateur de bénéfi-

cier de l'environnement virtuel visuel synthétisé par le dispositif. Le rendu sonore spatialisé restitué par le SMART-I², car il est basé sur une reconstruction physique du champ sonore, permet par contre à plusieurs utilisateurs de bénéficier simultanément de l'environnement virtuel audio synthétisé par le dispositif. Avec les moyens actuellement disponibles, une façon de proposer un rendu visuel monoscopique à deux utilisateurs consisterait à utiliser les filtres polarisants déjà en place pour envoyer aux deux yeux de chaque utilisateur l'image de la scène correspondant à leur position. La faisabilité et la qualité visuelle d'une solution technologique de type *monoscopie passive avec suivi* méritent donc d'être étudiées.

Chapitre 3

Applications de réalités virtuelles basées sur le SMART-I²

Sommaire

3.1	Introduction	57
3.2	GAVIP : “Gestural Auditory and Visual Interactive Platform”	58
3.2.1	Concept et motivations du projet	58
3.2.2	Architecture logicielle mise en place	59
3.2.3	Un premier scénario d’interaction : «Bubulles»	60
3.2.4	Vers un second scénario : «Lava Lova»	61
3.2.5	Perspectives	63
3.3	Intégration de MARC (Multi-modal Affective Reactive Characters) dans le SMART-I²	64
3.3.1	La plateforme MARC (Multi-modal Affective Reactive Characters)	64
3.3.2	Scène de démonstration développée pour les 5ièmes journées de l’Association Française de Réalité Virtuelle (AFRV)	65
3.3.3	Premier retour «utilisateur» à propos du rendu audio-visuel proposé par le SMART-I ²	66
3.3.4	Perspectives	68
3.4	Conclusion	68

3.1 Introduction

Le chapitre précédent a souligné l’impact que pouvait avoir l’ajout d’un système de *rendu sonore spatialisé* dans des applications de réalité virtuelle, en particulier sur la sensation de *présence*. L’importance du contenu, *i.e.* du monde virtuel qui est présenté au participant et des possibilités d’interactions qui lui sont offertes, a aussi été notée.

Dans ce chapitre sont décrits deux projets en cours (GAVIP, voir section 3.2, et MARC & SMART-I², voir section 3.3) dont l’objectif à long terme est l’étude de l’impact des différents attributs d’un *rendu sonore*

spatialisé sur la sensation de *présence* perçue. L'un des projets propose une immersion *active* des participants dans le monde virtuel (GAVIP), tandis que l'autre propose une immersion plutôt *passive* (MARC & SMART-I²). L'influence du contenu et de la tâche seront donc aussi potentiellement étudiés.

Étant donné la diversité des compétences requises pour atteindre ces objectifs, les deux projets sont le fruit de collaborations avec différents équipes du LIMSI, *Architectures et Modèles pour l'Interaction* (AMI) et *Acoustique et Audio* (A&A), ainsi qu'avec l'équipe *Interactions Musicales en Temps Réel* (IMTR) de l'IRCAM. Au sein de ces projets, ma principale contribution a été la création des outils permettant un contrôle aisée et flexible du *rendu sonore spatialisé* généré par le SMART-I² dans des contextes variés.

3.2 GAVIP : “*Gestural Auditory and Visual Interactive Platform*”

3.2.1 Concept et motivations du projet

La plateforme GAVIP¹ (*Gestural Auditory and Visual Interactive Platform*) a pour but l'étude des possibilités d'*interactions gestuelles* en temps réel entre un humain et des *objets virtuels audio-visuels et spatialisés*. L'*interaction gestuelle* est ici définie comme la relation bidirectionnelle artificiellement créée entre les gestes humains et les objets composant le monde virtuel. Les *objets virtuels audio-visuels et spatialisés* sont pensés comme des stimuli audio et visuels possédant une forte *cohérence inter-modale*, c'est à dire des stimuli audio-visuels qui sont cohérents spatialement, temporellement et sémantiquement. L'hypothèse sur laquelle repose ce projet est alors la suivante :

«*La perception et la qualité des interactions gestuelles sont améliorées par les cohérences spatiale, temporelle et sémantique entre les composantes audio et visuelles des objets composant le monde virtuel (cohérence inter-modale).*»

Sans avoir jamais été explicitée ni vraiment validée, cette hypothèse est pourtant partagée par des communautés scientifiques variées : *interactions “hommes-machines” (HCI)* [Vertegaal & Eaglestone 1998], *nouvelles interfaces pour l'expression musicale (NIME)* [Rodet et al. 2005] ou encore *réalité virtuelle* (voir section 2.2.1). La mise en relation *directe* des paramètres issus de l'analyse gestuelle et de ceux de synthèse sonore conduit généralement à des interactions assez pauvres. L'objectif est alors de créer des interactions plus riches en utilisant des descripteurs de gestes plus sophistiqués ou des relations *indirectes* entre les paramètres issus de l'analyse gestuelle et les paramètres de synthèse audio-visuelle [Wanderley & Depalle 2004].

C'est dans le but d'étudier expérimentalement l'hypothèse énoncée précédemment qu'a été créé le groupe de travail GAVIP, composé de doctorants de différents laboratoires : B. Caramiaux², S. Fdili Alaoui³, T. Bouchara & G. Parseihian⁴, M. Courgeon⁵ et moi-même. Dans ce groupe mon rôle a été de gérer les différents aspects relatifs à l'*espace* : génération des positions spatiales des objets virtuels (modèle statistique, physique) et spatialisation sonore (SMART-I² et WFS). Une architecture de rendu complète a

1. Le site web du projet est disponible à l'adresse : <http://imtr.ircam.fr/imtr/GAVIP>.

2. IRCAM, CNRS UMR 9912, Groupe “Interactions Musicales en Temps Réel”, 75004 Paris

3. IRCAM, CNRS UMR 9912, Groupe “Interactions Musicales en Temps Réel”, 75004 Paris & LIMSI, CNRS UMR 3251, Groupe “Architecture et modèles pour l'interaction”, 91403 Orsay

4. LIMSI, CNRS UMR 3251, Groupe “Acoustique et Audio”, 91403 Orsay

5. LIMSI, CNRS UMR 3251, Groupe “Architecture et modèles pour l'interaction”, 91403 Orsay

été mise en place et deux scénarios d’interaction ont été développés et ont fait l’objet d’une présentation lors de la conférence DAFX-11 à l’IRCAM [Caramiaux et al. 2011].

3.2.2 Architecture logicielle mise en place

La *cohérence inter-modale* entre les composantes sonores et visuelles des objets virtuels est capitale pour que les utilisateurs intègrent les flux audio et visuels en un unique *percept* [Spence 2007]. Cependant, dans les applications audio-visuelles interactives courantes, l’approche choisie est généralement de synthétiser séparément les flux audio et visuels. Des incohérences spatio-temporelles entre les deux flux existent alors, ce qui peut conduire à des dégradations de la perception et de la qualité des interactions (voir section 7.3). Pour minimiser ces effets indésirables, la *cohérence inter-modale* (audio, visuelle et spatiale) est recherchée, et l’architecture de rendu devra donc synthétiser simultanément les flux correspondant à chaque modalité et les informations spatiales qui y sont associées.

Pour répondre à cette exigence, l’architecture mise en place par GAVIP se base donc sur un unique bloc de *contrôle* pilotant le monde virtuel (voir figure 3.1). Ce bloc reçoit les résultats de l’*analyse gestuelle* et pilote ensuite les *synthèses* audio, visuelle, et spatiale de chacun des objets virtuels. Avoir un seul bloc de contrôle pour la modalité auditive, la modalité visuelle et l’espace permet donc a priori de maximiser la *cohérence inter-modale*.

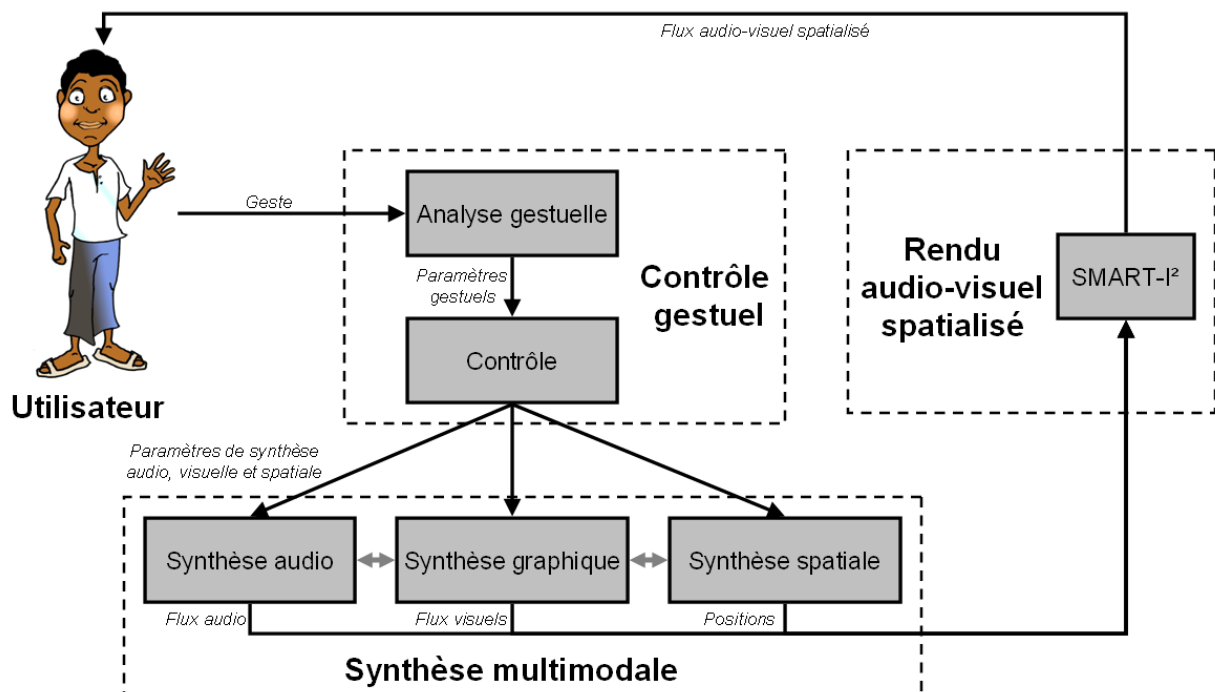


FIG. 3.1 – Vue d’ensemble de l’architecture logicielle et de rendu développée pour le projet GAVIP. La *cohérence inter-modale* est symbolisée par les doubles flèches grises présentes entre les différents modules de synthèse.

3.2.3 Un premier scénario d'interaction : «Bubulles»

Ce premier scénario d'interaction basé sur l'architecture décrite à la section 3.2.2 et nommé «Bubulles» a été conçu et présenté en tant que démonstration du SMART-I² à l'occasion de la conférence *Euro VR-EVE 2010*⁶ qui a eu lieu au LIMSI en mai 2010 [Rébillat et al. 2010a].

Dans ce scénario, le monde virtuel visuel est un sol plat et texturé au-dessus duquel évoluent des *bulles audio-visuelles* (voir figure 3.2). La composante visuelle de ces bulles est une sphère dont la texture est choisie parmi des jeux de textures classés par teintes. La composante sonore de ces bulles est un son court pré-enregistré. Des bruits d'impacts d'objets légers sur des surfaces rigides ont été choisis et répartis dans différentes banques de sons en fonction de leur provenance : goutte d'eau tombant sur le sol, grains de riz chutant sur une cithare ou sur une plaque métallique, etc. Ces *bulles audio-visuelles* ont une durée de vie dans le monde virtuel égale à la durée du son qui leur est associé. Dans le SMART-I², 16 *bulles audio-visuelles* peuvent être simultanément créées dans le monde virtuel.

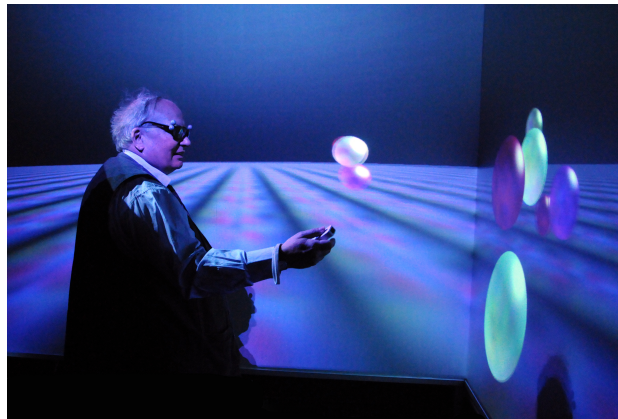


FIG. 3.2 – Un utilisateur interagissant avec le monde virtuel proposé par «Bubulles».

Du point de vue de l'*interaction gestuelle*, la difficulté était d'arriver à créer un lien entre les gestes de l'utilisateur et ce monde virtuel composé de *bulles audio-visuelles*. Étant donné le nombre important d'objets virtuels pouvant exister simultanément, un contrôle direct de chacun de ces objets n'était pas envisageable. Une approche statistique a alors été adoptée pour les blocs de *synthèse* (voir figure 3.1) et l'ensemble de ces *bulles audio-visuelles* a été traité comme un *nuage de bulles audio-visuelles*. Les paramètres à contrôler relativement à chacune des *bulles* ont ensuite été divisés en deux catégories : les paramètres *spatio-temporels* (distance, position, fréquence d'apparition) et les paramètres *sémantiques* (son joué, textures appliquées). Le choix a été fait de ne contrôler par le geste que les paramètres *spatio-temporels*. Les paramètres *sémantiques* sont tirés de façon aléatoire parmi une banque de sons et une palette de textures prédéfinies. Les paramètres *spatio-temporels* sont eux aussi tirés de façon aléatoire mais parmi une densité de probabilité dont les bornes sont contrôlables par le geste.

L'interaction gestuelle a été développée en utilisant les accéléromètres et les gyroscopes disponibles dans une Wiimote (manette de jeu de la console Nintendo Wii). Le bloc d'*analyse gestuelle* de la figure 3.1 extrait trois informations des données envoyées par les capteurs de la Wiimote : les angles de *rotation* et de *nutation* de la manette ainsi que la *période temporelle* à laquelle elle est «secouée». Le bloc de *contrôle*

6. <http://eurovr-eve-2010.limsi.fr/>

calcule ensuite les bornes des densités de probabilité associées au paramètres *spatio-temporels* en fonction de ces données :

- l’angle de *rotation* définit l’intervalle en azimuth dans lequel peut apparaître une bulle.
- l’angle de *nutation* définit l’intervalle en distance dans lequel peut apparaître une bulle.
- la *période temporelle* des secousses permet de déterminer l’intervalle temporel séparant deux bulles.

Ce type d’interaction gestuelle permet alors à l’utilisateur de positionner le *nuage de bulles audio-visuelles* où il veut dans l’espace en azimuth et en distance. Il peut aussi contrôler la densité de ce *nuage de bulles audio-visuelles* en allant de quelques gouttes éparées jusqu’à l’orage. Cependant, cette approche présente certains défauts. En particulier, le degré de contrôle de l’utilisateur, et donc l’expressivité offerte par l’interface sont assez limités. D’autre part, les *bulles audio-visuelles* ont un comportement discontinu assez basique (apparition, disparition) qui permet peu d’interactions à long terme.

3.2.4 Vers un second scénario : «Lava Lova»

Pour surmonter les limitations du scénario «*Bubulles*», un autre scénario nommé «*Lava Lova*» a été développé. Le premier objectif de ce nouveau scénario est de proposer aux utilisateurs des moyens plus doux et plus variés d’exercer un contrôle gestuel sur le monde virtuel. D’autre part, dans le scénario «*Bubulles*» aucune *cohérence inter-modale* n’existe au niveau *sémantique* (les bruits d’impact n’ont aucun lien avec les bulles visuelles). Le second objectif de «*Lava Lova*» est alors de proposer une *cohérence inter-modale* sémantique qui soit perceptible par les utilisateurs.

Monde virtuel

L’environnement visuel proposé à l’utilisateur est cette fois l’intérieur d’une *lampe à lave* (appelée encore *lampe à magma*, voir figure 3.3). Visuellement, les *bulles audio-visuelles* sont des *amas de magma visuel* se déformant et évoluant dans la lampe. À chaque *amas de magma* est associé un flux sonore fait de la concaténation de *grains sonores* obtenus en segmentant des enregistrements de sons environnementaux (vent dans les arbres, ruissellement, etc.).

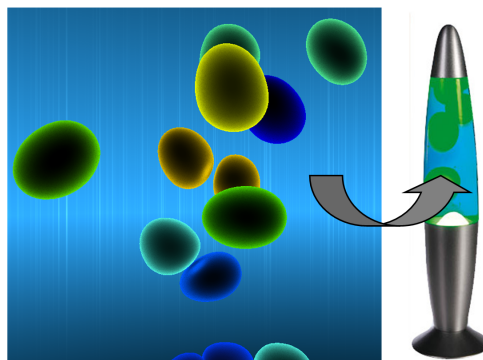


FIG. 3.3 – Schéma de principe du monde virtuel proposé par «*Lava Lova*». L’image de gauche est une capture d’écran du rendu visuel généré par le système.

Synthèse spatiale par modèle physique

Dans l'optique de proposer un contrôle gestuel qui soit à la fois plus doux et plus varié que dans le précédent scénario, le bloc de *synthèse spatiale* de la figure 3.1 a été doté d'un modèle physique calculant l'évolution temporelle des positions des $N = 16$ *amas de magma audio-visuels* à partir de conditions initiales données. Les interactions des N amas de magma entre eux, avec leur environnement, et avec l'extérieur sont gouvernées par la seconde loi de Newton. Chaque amas est assimilé à une masse ponctuelle m_n , électriquement chargée, et son accélération instantanée $\vec{a}_n(t)$ est donnée par :

$$m_n \vec{a}_n(t) = \sum_{k \neq n} \vec{F}_{k \rightarrow n}(t) + \vec{F}_{\text{centre} \rightarrow n}(t) + \vec{F}_{\text{ext} \rightarrow n}(t) \quad (3.1)$$

où

- $\sum_{k \neq n} \vec{F}_{k \rightarrow n}(t)$ est la somme de l'ensemble des forces appliquées par les autres amas sur le n -ième amas. Chaque force $\vec{F}_{k \rightarrow n}(t)$ est la somme d'une contribution visco-élastique et d'une contribution électrostatique.
- $\vec{F}_{\text{centre} \rightarrow n}(t)$ est une force centrale visco-élastique appliquée par l'environnement sur l'amas n .
- $\vec{F}_{\text{ext} \rightarrow n}(t)$ représente l'ensemble des forces externes à l'environnement et appliquées sur l'amas n . C'est la somme d'une force due à une rotation globale et de celle induite par un champ magnétique extérieur.

Ce modèle physique permet donc de calculer, à partir de conditions initiales données l'évolution des positions des 16 *amas de magma audio-visuel* en fonction du temps.

Synthèse sonore

La synthèse sonore est réalisée par l'intermédiaire de CataRT, développé par Diemo Schwarz à l'Ircam [Schwarz et al. 2006]. Ce logiciel prend en entrée une banque de sons et segmente chacun des sons présents en *grains sonores* de 242 ms. Il analyse ensuite chacun des grains et en extrait un certain nombre de descripteurs audio. À la suite de cette analyse, chaque grain est décrit par un vecteur (dont les coordonnées sont les descripteurs audio) et peut alors être positionné dans un espace multidimensionnel (où les dimensions correspondent aux descripteurs audio). De cette façon un espace bidimensionnel de *grains sonores* peut être défini en choisissant comme descripteurs audio le *niveau moyen* et le *centroïde spectral*. Lorsque le module de synthèse sonore reçoit une position dans l'espace 2D des *grains sonores*, il joue alors le grain le plus proche de la position indiquée.

Composante sémantique de la cohérence inter-modale

Il est alors très facile d'imaginer une relation *sémantique* entre les modalités audio et spatiale. En effet, les positions de chacun des *amas de magma audio-visuel* peuvent servir d'entrées au module de synthèse sonore. Alors les trajectoires *spatiales* des *amas de magma audio-visuels* seront aussi des trajectoires dans l'espace des *grains sonores*. Cela définit alors une certaine cohérence entre les aspects *sémantiques* liés à la modalité audio et ceux liés à la modalité *spatiale* (voir figure 3.4). Concrètement, le monde virtuel est

alors composé de 16 *amas de magma audio-visuel* naviguant de façon cohérente⁷ dans l’espace physique virtuel et dans l’espace timbral des *grains sonores*.

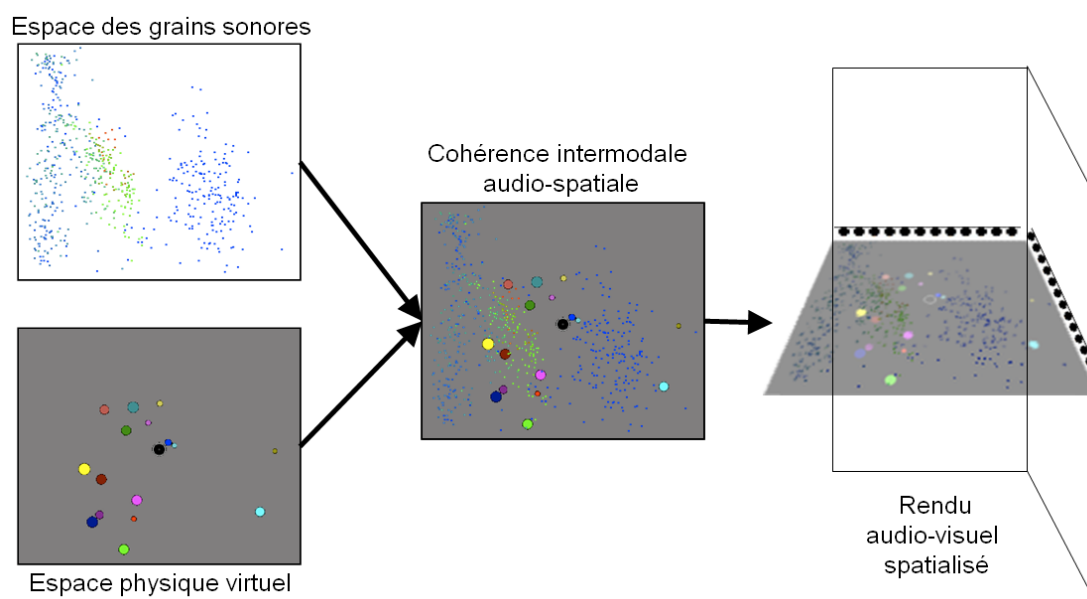


FIG. 3.4 – Illustration de l’ajout d’une composante *sémantique* à la cohérence inter-modale en faisant correspondre l’espace physique virtuel et l’espace timbral des *grains sonores* créée par le scénario «*Lava Lova*».

Contrôle gestuel

Actuellement, l’analyse gestuelle associée à ce scénario n’est pas encore fonctionnelle ni clairement établie. L’idée générale est cependant d’essayer de capter les mouvements de la personne de façon plus globale qu’avec simplement une Wiimote. D’un point de vue technologique, une caméra peut par exemple être utilisée. Alors le module d’*analyse gestuelle* pourrait chercher à estimer la *périodicité* et l’*énergie* présentes dans l’attitude globale du participant. Le comportement physique du système pourrait alors être modifié de façon à ce qu’un geste *énergique* avec une *périodicité* élevée induise une réduction de la viscosité dans le modèle physique et une augmentation des forces extérieures. À l’opposé un geste *tranquille* (avec une faible énergie et une faible périodicité) induirait un comportement du modèle physique qui tendrait à être stationnaire.

3.2.5 Perspectives

La plateforme développée par le projet GAVIP est donc fonctionnelle et deux scénarios ont été développés sur cette base. Au-delà du développement à finaliser pour le scénario «*Lava Lova*», la question de savoir dans quelle mesure *la perception et la qualité des interactions gestuelles sont améliorées par cohérence inter-modale des objets composant le monde virtuel* reste toujours ouverte. Il est envisageable de se

7. Une extension de la *cohérence inter-modale* à la modalité visuelle est aussi envisageable de façon similaire en créant un espace bidimensionnel de textures à appliquer sur les *amas de magma visuels*.

baser sur les scénarios existants pour réaliser des expériences perceptives visant à apporter des éléments de réponses à cette question. Ces scénarios peuvent aussi indirectement servir à l'étude de l'impact des différents attributs d'un *rendu sonore spatialisé* sur la sensation de *présence* (voir tableau 2.1).

3.3 Intégration de *MARC* (Multi-modal Affective Reactive Characters) dans le SMART-I²

Dans cette section sont présentés les travaux qui ont été menés en commun avec le groupe *Architecture et Modèles pour l'Interaction* (AMI) du LIMSI et plus spécifiquement avec M. Courgeon, C. Clavel et J.-C. Martin, ainsi qu'avec B. F.G. Katz du groupe *Acoustique et Audio* (AA) du LIMSI. Ces travaux visent à intégrer une plate-forme logicielle dédiée au rendu graphique de personnages virtuels réactifs et capables d'émotions au sein du SMART-I². Mon rôle a été de mettre en place une plateforme logicielle permettant un contrôle de la spatialisation sonore à distance ainsi que de rédiger la partie de l'évaluation relative à la *présence* et à la *cohérence audio-visuelle*. Cette intégration a donné lieu à la création d'une scène audio-visuelle qui a servi de base pour une démonstration et une présentation lors des 5ièmes journées de l'Association Française de Réalité Virtuelle (AFRV) [Courgeon et al. 2010]. L'objectif à long terme de cette collaboration est l'étude de l'impact des différents attributs d'un *rendu sonore spatialisé* sur la *perception des émotions* (voir section 2.3.3).

3.3.1 La plateforme *MARC* (Multi-modal Affective Reactive Characters)



FIG. 3.5 – Les quatre personnages virtuels capables d'exprimer des émotions proposés par la plate-forme *MARC* [Courgeon et al. 2008].

MARC (Multimodal Affective and Reactive Characters) est une plate-forme permettant la simulation en temps réel d'interactions émotionnelles avec un ou plusieurs personnages virtuels (voir figure 3.5). Les simulations proposées par cette plate-forme sont basées sur différents modèles émotionnels issus de la psychologie des émotions [Courgeon et al. 2008]. Les visages des personnages virtuels générés par *MARC* sont ainsi capables d'exprimer des émotions telles que la surprise, la colère, la peur, la compassion, etc.

Leur posture générale (*i.e.* la position de leurs mains, de leurs bras, etc.) est de plus contrôlable en temps réel et participe directement à l'émotion perçue.

Le rendu graphique généré en temps réel par la plateforme *MARC* est contrôlable par des applications externes en utilisant des protocoles standards de communication. Une banque d'animations prédéfinies (visage et posture) correspondant à un certain nombre d'émotions et d'animations est par ailleurs disponible ce qui permet de créer aisément des scènes dynamiques et interactives. À l'heure actuelle, quatre personnages différents sont disponibles.

3.3.2 Scène de démonstration développée pour les 5^{èmes} journées de l'Association Française de Réalité Virtuelle (AFRV)

Dans le but d'amorcer la collaboration avec l'équipe AMI, une première phase de développement consistait en l'intégration des personnages proposés par la plate-forme *MARC* dans le monde virtuel audio-visuel généré par le SMART-I². La figure 3.6 montre le résultat de cette intégration⁸.



FIG. 3.6 – Un utilisateur interagissant avec les personnages virtuels proposés par la plate-forme *MARC* qui sont intégrés dans le SMART-I².

Ensuite, ces personnages ont été animés dans le but de créer une courte scène audio-visuelle et interactive faisant office de démonstration lors des 5^{èmes} journées de l'Association Française de Réalité Virtuelle (AFRV). Dans ce but, les mouvements et les paroles de deux acteurs amateurs improvisant une courte scène ont été enregistrés. Ces données ont ensuite servi de base à l'animation gestuelle des personnages virtuels et à la création des flux sonores associés à chacun d'eux. Le scénario simpliste de la démonstration proposée se résume alors ainsi :

«John, photographe, désire faire des clichés d'une danseuse. Sarah, la danseuse, s'efforce de répondre à ses attentes. John est très exigeant et Sarah s'impatiente. N'obtenant pas de bons résultats avec elle, John s'adresse alors à l'utilisateur et lui demande s'il souhaite apparaître sur les clichés à la place de Sarah. Il positionne ensuite l'utilisateur dans l'espace et le photographie en le complimentant de son talent. La

8. Un extrait de la scène de démonstration est visionnable à l'adresse <http://www.youtube.com/watch?v=TppB4V4GXGw>.

position de l'utilisateur dans la zone de rendu étant connue, John le fixe du regard durant la seconde partie.»

3.3.3 Premier retour «utilisateur» à propos du rendu audio-visuel proposé par le SMART-I²

Lors des 5ièmes journées de l'Association Française de Réalité Virtuelle (AFRV), les participants avaient la possibilité d'assister individuellement à la scène de démonstration décrite à la section 3.3.2. À la suite de leur expérience dans le monde virtuel, il leur était demandé de noter sur une échelle allant de 0 (pas du tout d'accord) à 4 (complètement d'accord) leurs impressions par rapport au rendu audio-visuel spatialisé. Le questionnaire à remplir est composé d'une impression sur la *cohérence spatiale audio-visuelle* et de quatre impressions concernant la sensation de *présence* (voir tableau 3.1). Ces phrases sont reprises de questionnaires de *présence* établis comme sensibles aux aspects spatiaux du rendu sonore [Bormann 2005].

1. *J'ai eu l'impression que le son et l'image étaient spatialement cohérents.*
2. *J'ai eu l'impression d'être face à deux personnes dans un studio de photographie et non de regarder une animation enregistrée.*
3. *J'oubliais le monde réel et le monde virtuel devenait réel pour moi.*
4. *Si «0» signifie que pendant la démonstration j'étais complètement immergé dans le monde réel et «4» que j'étais complètement immergé dans le monde virtuel, quelle est ma sensation d'être immergé dans le monde réel ou dans le monde virtuel ?*
5. *Je me souviens du studio du photographe plutôt comme une image vue sur un écran d'ordinateur (0) que comme un endroit que j'ai visité (4).*

TAB. 3.1 – Questionnaire proposé aux participants à la suite d'une expérience dans le monde virtuel. Des réponses allant de 0 (pas du tout d'accord) à 4 (complètement d'accord) pouvaient être données à chacune des questions.

Ce questionnaire a été rempli par 16 participants à la suite de leur immersion dans le monde virtuel audio-visuel proposé par la scène de la section 3.3.2. Ces 16 participants ne forment pas un échantillon *représentatif* de la population étant donné qu'ils sont tous plus ou moins spécialistes des aspects graphiques liés la réalité virtuelle. Les résultats concernant la *cohérence audio-visuelle* (question 1 du tableau 3.1) et concernant la *présence* (questions de 2 à 5 du tableau 3.1) ont été traités en deux blocs séparés. Les notes de la question 1 sont interprétées comme une traduction de la *cohérence audio-visuelle* ressentie. La somme des notes des questions 2 à 5 donne une traduction de la sensation de *présence* ressentie. Les résultats des 16 participants concernant la *cohérence audio-visuelle* et la *présence* sont représentés sous forme d'histogrammes à la figure 3.7.

L'étude de la figure 3.7(a) révèle que tous les participants ont notés la *cohérence spatiale audio-visuelle* entre 2 et 4 avec une note moyenne de 3.2. Globalement, le rendu spatialisé audio-visuel restitué par le SMART-I² a été jugé comme procurant spatialement une bonne cohérence perceptive entre les modalités audio et visuelle. Cela est important car d'éventuelles incohérences spatiales entre les modalités audio et visuelle peuvent nuire à la qualité audio-visuelle globale du rendu (voir section 2.3.5) et cela ne semble pas

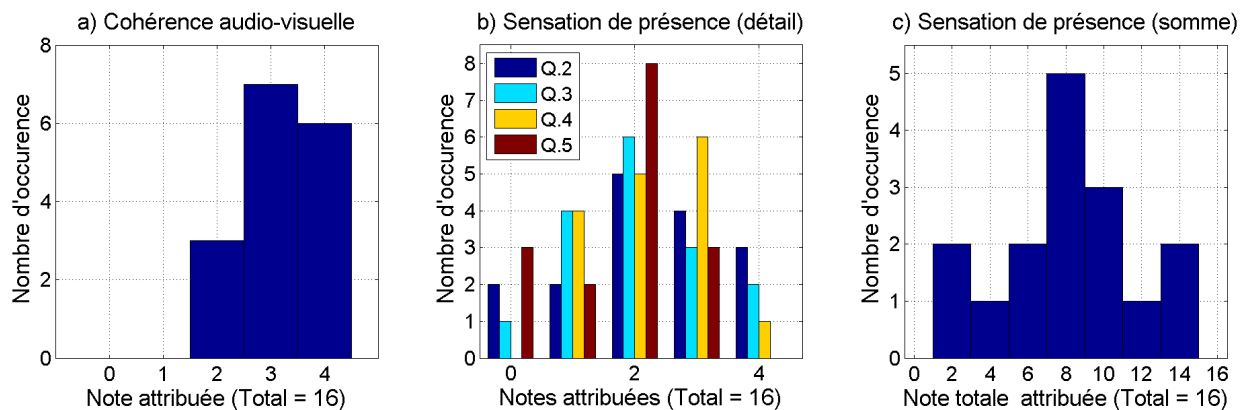


FIG. 3.7 – Réponses des participants de l'AFRV au questionnaire du tableau 3.1.

- a) Histogramme des notes des participants à la question 1 du tableau 3.1.
- b) Histogrammes des notes de chaque participant aux questions de 2 à 5 du tableau 3.1.
- c) Histogramme de la somme des notes de chaque participant aux questions de 2 à 5 du tableau 3.1.

être le cas ici.

L'étude des figures 3.7(b) et 3.7(c) montre par ailleurs que les participants ont ressentis la sensation de *présence* de façon très variable (toute l'échelle utilisée pour les questions de 2 à 4 et des notes globales allant de 2 à 14). La note totale moyenne attribuée à la *présence* est de 8.4, soit juste au dessus de la moyenne (qui est de 8 ici). Cela montre que le sentiment de *présence* dépend assez fortement des attentes de chaque participant par rapport au contenu audio-visuel et illustre aussi la difficulté de mesurer cette sensation (voir sections 2.2.2 et 2.2.4). Il est néanmoins satisfaisant de constater qu'en moyenne les participants se sont sentis correctement immergés dans le monde virtuel audio-visuel proposé par le SMART-I².

La validité *absolue* de ces résultats est cependant discutable, et dans les expériences futures, il serait plus judicieux d'utiliser une telle échelle dans le but de *comparer* les résultats que donnent différentes situations.

Les gens avaient par ailleurs l'opportunité de laisser des commentaires écrits relatant leur expérience dans le monde virtuel. Dans ces commentaires, les trois points suivants sont abordés de façon récurrente :

- *Le problème de l'intimité* : plusieurs personnes indiquent que la présence d'expérimentateurs dans leur dos durant la scène les a fait se sentir «*mal à l'aise*» et que «*plus d'intimité*» leur aurait davantage permis de ressentir une sensation de *présence* plus intense. L'autonomie du participant dans la scène audio-visuelle semble donc aider à se sentir *présent* dans le monde virtuel.
- *L'interaction entre le personnage virtuel et l'utilisateur favorise la présence* : en regardant les réactions des participants durant la scène, il était flagrant que certaines personnes répondaient verbalement au photographe virtuel et réagissaient spontanément à ses injonctions en se déplaçant dans le monde réel. Un participant a même avoué avoir «*éprouvé la sensation d'avoir satisfait le photographe en [se] déplaçant, comme s'il était réel*». Le scénario, et particulièrement le fait de prendre à parti le participant, joue donc beaucoup dans la sensation de *présence* ressentie par les utilisateurs. Cela est donc en accord avec l'importance du contenu soulignée dans la section 2.2.4.
- *La qualité audio-visuelle globale* : les participants étant majoritairement des spécialistes du rendu graphique, ils ont été particulièrement sensibles à la qualité de l'interface visuelle, qui n'est pas parfaite. Cependant l'un des participants rapporte que «*la cohérence du son et le scénario font oublier les imprécisions graphiques*». Cela fait directement écho aux interactions entre les qualités audio et visuelle

évoquées à la section 2.3.5.

Ce premier retour des utilisateurs sur le rendu audio-visuel spatialisé restitué par le SMART-I² est donc assez positif. Les participants ont perçu une très bonne cohérence audio-visuelle et ont expérimenté un sentiment de *présence* variable, mais réel. Les commentaires font par ailleurs apparaître que le scénario, et en particulier le fait de prendre à parti le participant, participe sans doute de façon très importante à la sensation de *présence*. On peut donc formuler l’hypothèse que la cohérence audio-visuelle et l’interaction avec le personnage virtuel sont ici les facteurs qui ont permis d’augmenter la sensation de *présence* perçue.

3.3.4 Perspectives

Comme évoqué dans les sections 2.3.2 et 2.3.3, l’ajout d’un *rendu sonore spatialisé* peut avoir un impact fort sur la sensation de *présence* qu’ont les utilisateurs des personnages virtuels et sur la perception des *émotions* qu’ils expriment. Cependant, comme le montre le tableau de synthèse 2.1 de la section 2.3.6, la nature de ces impacts reste encore mal connue, voire même inconnue dans certains cas. Les premières réactions des utilisateurs suggèrent que l’intégration de la plateforme *MARC* au sein du SMART-I² peut fournir une bonne base pour l’étude de ces questions fondamentales.

3.4 Conclusion

Dans ce chapitre sont décrits deux projets en cours (GAVIP, voir section 3.2, et MARC & SMART-I², voir section 3.3) dont l’objectif à long terme est l’étude de l’impact des différents attributs d’un *rendu sonore spatialisé* sur la sensation de *présence* perçue. L’un des projets propose une immersion *active* des participants dans le monde virtuel (GAVIP), tandis que l’autre projet propose une immersion plutôt *passive* (MARC & SMART-I²). Ces projets permettent donc aussi potentiellement l’étude de l’influence du contenu et de la tâche sur la sensation de *présence*.

Le projet GAVIP n’a pas encore été évalué par les utilisateurs et ne fournit donc pas encore de résultats relativement aux objectifs mentionnés précédemment. Les deux scénarios développés semblent cependant prometteurs. Le projet MARC & SMART-I² a permis une évaluation assez sommaire qui permet d’émettre l’hypothèse que la cohérence audio-visuelle et l’interaction avec le personnage virtuel ont permis d’augmenter la sensation de *présence* perçue. Ces deux projets ouvrent donc des perspectives nombreuses et intéressantes dans le domaine des *interfaces “hommes/machines”*.

Chapitre 4

Mesure des non-linéarités dans les systèmes vibrants

Sommaire

4.1	Introduction	69
4.2	Identification de modèles de Hammerstein en cascade pour la description des non-linéarités dans les systèmes vibrants	70
4.2.1	Introduction	71
4.2.2	Modélisation et mesure des non-linéarités	72
4.2.3	Bases mathématiques de la méthode	75
4.2.4	Implémentation pratique	78
4.2.5	Validation de la méthode	80
4.2.6	Modélisation de transducteurs acoustiques par des modèles de Hammerstein en cascade	84
4.2.7	Prédiction de la distorsion harmonique générée par les transducteurs	87
4.2.8	Discussion	91
4.2.9	Conclusion	91
4.3	Discussion	93
4.3.1	Effets de différents paramètres sur la qualité de l'estimation des noyaux	93
4.3.2	Comparaison avec les travaux d'A. Novák <i>et al.</i>	95
4.3.3	Chaîne de systèmes non-linéaires	101
4.3.4	Librairie Matlab <i>Hammerstein Toolbox</i>	107
4.4	Conclusion et perspectives	107

4.1 Introduction

Beaucoup d'études menées sur les systèmes vibrants supposent ces derniers *linéaires* sans toujours le vérifier. Or, les systèmes vibrants courants (haut-parleurs, pots vibrants, etc.) sont en réalité au moins faiblement non-linéaires et ces composantes non-linéaires peuvent affecter significativement les mesures

effectuées. La problématique générale de ce chapitre a trait à l'estimation et éventuellement à la correction des non-linéarités générées par les systèmes faiblement non-linéaires.

Ce chapitre est organisé en deux parties distinctes. La première partie (section 4.2) présente un article intitulé "*Identification of cascade of Hammerstein models for the description of non-linearities in vibrating devices*" paru dans le volume 330(5) du *Journal of Sound and Vibration* et rédigé en anglais en collaboration avec Romain Hennequin. Cet article présente une méthode originale, dérivée des travaux d'Angelo Farina [Farina 2000, Farina et al. 2001], permettant d'estimer rapidement un modèle de système non-linéaire correspondant à un système vibrant. Romain Hennequin a contribué de façon importante à la formulation mathématique de la méthode (voir section 4.2.3) ainsi qu'à son implémentation (voir section 4.2.4). La seconde partie de ce chapitre (section 4.3) présente une discussion autour de cette nouvelle méthode. Dans un premier temps l'influence de différents paramètres sur la qualité des noyaux estimés est discutée (section 4.3.1). Puis, dans la section 4.3.2, la méthode développée est comparée aux travaux d'Antonin Novák qui ont été menés en parallèle au Laboratoire d'Acoustique de l'Université du Maine (LAUM) sur une thématique très proche. Dans un troisième temps (section 4.3.3) le cas d'une chaîne de systèmes non-linéaires est étudié et les possibilités offertes par la méthode développée ici sont discutées. Enfin en section 4.3.4, la librairie Matlab disponible en ligne et proposant une implémentation des méthodes proposées en section 4.2 et par Novák et al. [2010a] est brièvement introduit.

4.2 Identification de modèles de Hammerstein en cascade pour la description des non-linéarités dans les systèmes vibrants

Marc Rébillat, Romain Hennequin¹, Étienne Corteel & Brian F.G. Katz

Cette section reproduit l'article intitulé "*Identification of cascade of Hammerstein models for the description of non-linearities in vibrating devices*" paru dans le volume 330, numéro 5, pages 1018-1038 du *Journal of Sound and Vibration* [Rébillat et al. 2011c]. Une partie de cet article a fait l'objet d'une présentation lors de la 128^{ème} convention de l'*Audio Engineering Society* à Londres [Rébillat et al. 2010b].

Abstract

In a number of vibration applications, systems under study are slightly non-linear. It is thus of great importance to have a way to model and to measure these non-linearities in the frequency range of use. Cascade of Hammerstein models conveniently allows one to describe a large class of non-linearities. A simple method based on a phase property of exponential sine sweeps is proposed to identify the structural elements of such a model from only one measured response of the system. Mathematical foundations and practical implementation of the method are discussed. The method is afterwards validated on simulated and real systems. Vibrating devices such as acoustical transducers are well approximated by cascade of Hammerstein models. The harmonic distortion generated by those transducers can be predicted by the model over the entire audio frequency range for any desired input amplitude. Agreement with more time consuming classical distortion measurement methods was found to be good.

1. Institut TELECOM (TELECOM ParisTech, CNRS LTCI), 75013 Paris, France

4.2.1 Introduction

Vibratory phenomena are usually assumed to be linear. However, many vibrating systems are subject to non-linear behaviours, such as loudspeakers [Quaegebeur & Chaigne 2008], musical instruments [Gilbert et al. 2008] and vibrating plates [Ribeiro & Petyt 2000]. Even wave propagation in air is not completely linear [Gee et al. 2006]. The study of these non-linearities is thus of great importance in order to model these devices and phenomena or to justify their “linearity”.

Identification of non-linear systems requires measurements or estimation of model’s structural elements from a finite set of input/output data [Chen 1995]. Classical linear measurement methods [Muller & Massarani 2001, Stan et al. 2002] capture only the linear behaviour of the system under study. Traditional non-linear measurement methods [Czerwinski et al. 2001] give total harmonic distortion (THD), harmonic distortion of order n (HD_n) or inter-modulation products (IMP). These quantities are measured using pure tones at a given amplitude and frequency. They do not describe non-linearities themselves but only some of their effects for arbitrary excitations. Moreover, experimental processes involved in those methods are very time-consuming if a wide range of frequencies and amplitudes is to be considered. There is thus a real need for rapid model based procedures to measure non-linearities.

Non-linear systems can be classically represented by Volterra series [Hasler 1999] or by “Sandwich” structures [Chen 1995]. The cascade of Hammerstein models [Gallman 1975] is a subclass of those models and can be used to exactly represent systems having diagonal Volterra Kernels. This model is composed of N branches in parallel. Each branch comprised a static polynomial non-linearity followed by a linear filter.

A simple method which makes it possible to quickly access the structural elements of a cascade of Hammerstein models is presented in this paper. This method is based on the method initially proposed by Farina [Farina 2000]. Exponential sine sweeps are used as input signals, and allow for the temporal separation of the different orders of non-linearity [Farina 2000, Craven & Gerzon 1992, Norcross & Vanderkooy 1995, Griesinger 1996]. Structural elements of the assumed model are then extracted from only one response of the system. The method is validated on a simulated system and the influence of the different parameters is shown.

As transducers are most often the least linear part in the audio chain, knowing their non-linear behaviour is very important. Two major types of transducers exists: electrodynamic loudspeakers [Klippel 2006] and panel-based transducers (distributed mode loudspeakers [Harris & Hawksford 2000] or multi-actuator panels [Kuster et al. 2006]). In electrodynamic loudspeakers, a motor converts the electrical signal into motion and makes a cone vibrate. The piston-like movement of the cone generates the sound field. In panel-based loudspeakers, a motor is also used but transmits its motion to a light and stiff panel. The flexural waves travelling in the panel then generate the sound field.

In both loudspeaker types, the motor induces non-linearities because of non-uniform magnetic field, Eddy currents and variations of the electrical inductance with displacements [Klippel 2006, Colloms et al. 1998]. In electrodynamic loudspeakers, significant excursion can induce non-linear bending in the cone and a non-linear behaviour of the suspensions [Quaegebeur & Chaigne 2008]. In panel-based loudspeakers, large amplitude displacements occur in the plate near the exciter position. In this case the propagation of flexural waves [Ribeiro & Petyt 2000] and the strain/stress relation of the material which compose the plate [Daniel & Abot 2000] can be non-linear.

In the literature, electrodynamic loudspeakers have been greatly studied from a non-linear point of view. To represent their non-linear behaviour, different physical models have been built. Their formulation was either completely analytical [Kaizer 1987, Klippel 1996a] or based on the finite element method [Tsuchiya et al. 2003]. In Ref. [Klippel 1996b], Klippel proposed to reduce the Volterra series expansion to a “Sandwich” model and identified its parameters from measurements using the method presented in [Baumgartner & Rugh 1975]. In [Farina 2000, Hawksford 2005], it is suggested to use simpler models, *i.e.* cascade of Hammerstein models, to model and identify different audio systems, including acoustical transducers.

In the present paper, two different acoustical transducers (an electrodynamic one and a panel-based one) are studied experimentally under the assumption that they can be modelled using cascade of Hammerstein models. Their models are completely identified using the previously introduced method. THD and HD_n at different frequencies and amplitudes of the input signal are evaluated for these transducers by classical means and compared to predictions made using the identified Kernels. The agreement between the results given by the two methods is very good for a wide range of amplitudes.

After reviewing how to model and to measure non-linearities (section 4.2.2), the mathematical foundations of the current method are presented (section 4.2.3). Implementation of the method is then described (section 4.2.4) and validated on a simulated system (section 4.2.5). Acoustical transducers are identified using the previous method (section 4.2.6) and the resulting Kernels are used successfully to predict the harmonic distortion generated by the two transducers (section 4.2.7).

4.2.2 Modelling and measurements of non-linearities

An overview of existing models and measurement methods of non-linearities is given in this section. Only Single-Input/Single-Output (SISO) time-invariant causal non-linear systems without continuous component will be considered here.

Volterra series

Volterra series [Hasler 1999] enables one to express the relationship between the non-linear system input $e(t)$ and output $s(t)$ as a series of multiple convolution integrals :

$$s(t) = \sum_{k=1}^{+\infty} \int_0^{+\infty} \dots \int_0^{+\infty} v_k(\tau_1, \dots, \tau_k) e(t - \tau_1) \dots e(t - \tau_k) d\tau_1 \dots d\tau_k \quad (4.1)$$

The functions $\{v_k(t_1, \dots, t_k)\}_{k \in \mathbb{N}^*}$ are called Volterra Kernels and completely characterize the system. Volterra models can then be seen as a generalization of the simple convolution operator used for linear systems. Such models represent exactly any non-linear “analytical” system [Palm 1978], and approximate any non-linear system with a “fading memory” [Boyd & Chua 1985]. Measurement methods exists to identify the first two or three terms of Volterra series [Boyd et al. 1983, Bicken et al. 2002, Prazenica & Kurdila 2004, Kayran & Eksioglu 2005]. These experimental methods are time consuming because they require many measurements. Moreover the difficult physical interpretation of the different terms of the Volterra series limits its use [Voishvillo et al. 2004].

Sandwich approach

Another approach to non-linear system identification is to assume that systems have a given block-structure. Following the “Sandwich” approach [Chen 1995], a non-linear system is represented as M parallel branches composed of three elements in series: a static non-linear part $P_m(\cdot)$ sandwiched between two linear parts $l_m^l(t)$ and $l_m^r(t)$. Such systems are a subclass of Volterra systems. It can be shown that any continuous non-linear system can be approximated by such a model [Palm 1979].

To identify such structures, the form of the static non-linear part can be assumed and the two unknown linear parts can be estimated from measurements at different frequencies and amplitudes [Baumgartner & Rugh 1975]. This leads to only a discrete knowledge of $L_m^l(f)$ and $L_m^r(f)$ in the frequency domain and remains a long experimental task. In Ref. [Abel & Berners 2006], Abel *et al.* proposed another method to identify such structures. Unfortunately, this general method cannot be used successfully in practice due to numerical instabilities.

Cascade of Hammerstein models

In a cascade of Hammerstein models [Gallman 1975], each branch is composed of one non-linear static polynomial element followed by a linear one $h_n(t)$, as shown in Fig. 4.1.

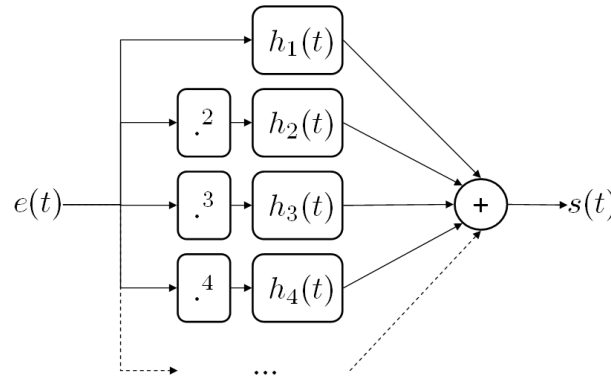


Figure 4.1: Block diagram representation of a cascade of Hammerstein models.

Mathematically, the relation between the input $e(t)$ and the output $s(t)$ of such a system is given by Eq. (4.2), where $*$ denotes the convolution.

$$s(t) = \sum_{n=1}^N h_n * e^n(t) \quad (4.2)$$

In this model, each impulse response $h_n(t)$ is convolved with the input signal elevated to its n^{th} power and the output $s(t)$ is the sum of these convolutions. The first impulse response $h_1(t)$ represents the linear response of the system. The other impulse responses $\{h_n(t)\}_{n \in \{2 \dots N\}}$ model the non-linearities.

The family $\{h_n(t)\}_{n \in \{1 \dots N\}}$ will be referred to as the Kernels of the model. These Kernels are assumed to be integrable. Any cascade of Hammerstein models is fully represented by its Kernels.

It can easily be shown from Eqs. (4.1) and (4.2) that cascade of Hammerstein models correspond to Volterra models having diagonal Kernels in the temporal domain, as in Eq. (4.3), where $\delta(t)$ represents

the Dirac distribution. This non-linear model is thus referred in the literature as a diagonal Volterra model [Farina et al. 2001], but also as a cascade of Hammerstein models [Chen 1995] or Uryson model [Gallman 1975].

$$\forall(\tau_1, \dots, \tau_k) \quad v_k(\tau_1, \dots, \tau_k) = h_k(\tau_1)\delta(\tau_1) \dots \delta(\tau_k) \quad (4.3)$$

As can be seen in Eq. (4.2), cascade of Hammerstein models are linear in the parameters to be estimated, *i.e.* the output of the system is a linear combination of the Kernels $\{h_n(t)\}_{n \in \{1 \dots N\}}$. A naive approach is to identify the model using a classical least square method, as proposed for general Volterra systems in [Reed & Hawksford 1996]. Thus the mean squared error between the actual output of the system $y(t)$ and the output of the estimated model $s(t)$ given in Eq. (4.2) can be minimized with respect to the coefficients of $h_1(t), h_2(t), \dots, h_N(t)$ and the solution is given by:

$$\underset{h_1(t), h_2(t), \dots, h_N(t)}{\operatorname{argmin}} \sum_t \|y(t) - s(t)\|^2 \quad (4.4)$$

However, the least square method requires the inversion of a $MN \times MN$ matrix, where N is the order of the system under test and M is the length of the impulse responses $h_n(t)$ in samples. This matrix can be very ill-conditioned since it is generated from the exponent (until N) of the input signal. This results in important errors in parameters estimation especially in noisy conditions. Moreover the computation of the matrix from the input signal and of the inverse of the matrix is computationally costly and limits in practical case the memory M of the system. Some numerical methods are however available to limit these points (see for example [Golub 6 29, Chen et al. 1989]), and to overcome these drawbacks alternative methods have also been developed.

Gallman [Gallman 1975] and Hawksford [Hawksford 2005] proposed a method to estimate the elements of a cascade of Hammerstein models using Gaussian noise at different amplitudes as inputs. The employed estimation procedures are strongly based on the knowledge of the order of non-linearity of the polynomial expansion, which is unknown in practical cases. Moreover, these methods to identify the Kernels from the measurements are also computationally costly.

Farina proposed another method using sine sweeps with frequency varying exponentially with time [Farina 2000]. An upper bound of the order of non-linearity of the model has to be assumed. This method allows only for the separation of the different orders of non-linearity and not for the complete identification of the Kernels of the system. Recently, Novák *et al.* [Novák et al. 2009] have identified Kernels from the contributions of the different orders of non-linearity using a least mean square minimization procedure. No results are provided to judge the influence of the different parameters on its performances.

The method proposed here gives direct mathematical access to all the Kernels $\{h_n(t)\}_{n \in \{1 \dots N\}}$ from the contributions of the different orders of non-linearity obtained as in Ref. [Farina 2000]. The main advantage of the proposed approach over the least squared based technique, besides conditioning and computational problems, is that it provides a direct evaluation of the N impulse responses $h_n(t)$ of the system. The foundations and the key implementation of this simple method are explained in detail. The whole procedure is validated on a simulated system and on two real systems.

4.2.3 Mathematical foundations of the method

Mathematical foundations of the method used for direct identification of the elements of a cascade of Hammerstein models are given in this section. This is based on the procedure initially proposed by Farina [Farina 2000].

A cascade of Hammerstein models fed with sine sweeps

To experimentally cover the frequency range on which the system under study is to be identified, cosines with time-varying frequencies are interesting signals. Eq. (4.5) defines such a signal.

$$\forall t \in \mathbb{R} \quad e(t) = \cos(\Phi(t)) \quad (4.5)$$

If $e(t)$ is the input of the cascade of Hammerstein models, the output of the non-linear block of the i^{th} branch will have the form of Eq. (4.6), as can be seen in Fig. 4.1.

$$e^i(t) = \cos^i[\Phi(t)] \quad (4.6)$$

Using Chebyshev polynomials, $e^i(t)$ is rewritten in Eq. (4.7) as a linear function of $\{\cos[k\Phi(t)]\}_{k \in [1, i]}$. Details of the computation of the matrix C are provided in the appendix A.

$$\forall i \in \{1 \dots N\} \quad \cos^i[\Phi(t)] = \sum_{k=0}^i C(i, k) \cos[k\Phi(t)] \quad (4.7)$$

Exponential sine sweeps

When the instantaneous frequency of $e(t)$ is increasing exponentially from f_1 to f_2 ($f_1, f_2 > 0$) in a time T , such a signal is referred to as an “exponential sine sweep” [Farina 2000, Norcross & Vanderkooy 1995] and its instantaneous phase is given by :

$$\forall t \in \mathbb{R} \quad \Phi(t) = 2\pi \frac{f_1 T}{\ln \frac{f_2}{f_1}} (e^{\frac{t}{T} \ln \frac{f_2}{f_1}} - 1) - \pi/2 \quad (4.8)$$

The corresponding instantaneous frequency of $e(t)$ is :

$$\forall t \in \mathbb{R} \quad f(t) = \frac{\Phi'(t)}{2\pi} = f_1 e^{\frac{t}{T} \ln \frac{f_2}{f_1}} \quad (4.9)$$

Thus $f(0) = f_1$ and $f(T) = f_2$. The frequency range $[f_1, f_2]$ corresponds to the band of interest of the system under test.

Fundamental phase property

From Eq. (4.8), it can be shown that this type of signal exhibits the following phase property :

$$\forall k \in \mathbb{N}^*, \forall t \in \mathbb{R} \quad k\Phi(t) = \Phi(t + \frac{T \ln k}{\ln \frac{f_2}{f_1}}) - (k-1)(\frac{\pi}{2} + 2\pi \frac{f_1 T}{\ln \frac{f_2}{f_1}}) \quad (4.10)$$

By choosing $T_m = (2m\pi - \frac{\pi}{2}) \frac{\ln f_2/f_1}{2\pi f_1}$ with $m \in \mathbb{N}^*$, the second term in Eq. (4.7) becomes a multiple of 2π and one obtains Eq. (4.11) which is another way to express the k^{th} term of the linearisation presented in Eq. (4.7).

$$\forall k \in \mathbb{N}^*, \quad \cos(k\Phi(t)) = \cos(\Phi(t + \Delta t_k)) \quad \text{with} \quad \Delta t_k = \frac{T_m \ln k}{\ln(f_2/f_1)} \quad (4.11)$$

For any T_m -long logarithmic sweep, multiplying the phase by a factor k results in the same signal, but in advance in the time domain by Δt_k . As can be seen from Eq. (4.11), this time advance depends only on the sweep parameters T_m, f_1, f_2 and on k . In Refs. [Farina 2000, Morise et al. 2007], similar time advances were obtained using different arguments.

The fact that T_m must take only a discrete set of values to ensure the fundamental phase property Eq. (4.11) has been first shown in Ref. [Novák et al. 2009] but is mathematically demonstrated here for the first time.

One should note that $e(t)$ has been designed for all t with its instantaneous frequency increasing from f_1 to f_2 between $t = 0$ and $t = T$. In practice, signals are defined only on $[0, T]$. Thus the phase property is not valid on the whole support of the function. The phase property becomes false when the instantaneous frequency of $\cos[k\Phi(t)]$ is outside the frequency range of interest (i.e. $[f_1, f_2]$).

Inverse convolution

Using the Eqs. (4.11) and (4.2), one obtains :

$$s(t) = \sum_{n=1}^N g_n * e(t + \Delta t_n) + K \quad (4.12)$$

with:

$$g_n(t) = \sum_{k=1}^N C(k, n) h_k(t) \quad \text{and} \quad K = \sum_{n=1}^N C(n, 0) \int_{-\infty}^{+\infty} h_n(t) dt \quad (4.13)$$

In Eq. (4.13), $g_n(t)$ corresponds to the contribution of the different Kernels to the n^{th} harmonic. K is the global continuous component resulting from the continuous components of the different Kernels. As the Kernels are assumed to be integrable, K is correctly defined.

In order to separately identify each Kernel of the cascade of Hammerstein models, a signal $y(t)$ which looks like an inverse in the convolution sense of $e(t)$ is needed. Unfortunately, such an inverse does not necessarily exist mathematically. However, a band-limited inverse $y(t)$ can easily be defined such that it satisfies the relation (4.14) with $\text{sinc}(x) = \sin(\pi x)/\pi x$.

$$y * e(t) = \text{sinc}(2f_2 t) - \text{sinc}(2f_1 t) = d(t) \quad (4.14)$$

$d(t)$ can be seen as a band-limited Dirac Function, since its Fourier transform is $\mathbb{1}_{[-f_2, -f_1] \cap [f_1, f_2]}(f)$. Then, $Y(f)$, the Fourier transform of the inverse filter $y(t)$ can be built in the frequency domain using Eq. (4.15), where $E^*(f)$ is the complex conjugate of $E(f)$.

$$Y(f) = \frac{1}{E(f)} \mathbb{1}_{[-f_2, -f_1] \cap [f_1, f_2]}(f) \simeq \frac{E^*(f)}{|E(f)|^2 + \epsilon(f)} \quad (4.15)$$

In practice, the filter $Y(f)$ should be built by replacing the discontinuous function $\mathbb{1}_{[-f_2, -f_1] \cap [f_1, f_2]}(f)$ by a function which ensures a smoother transition between the two frequency domains and thus generates less unwanted side effects in the time domain.

$\epsilon(f) = \beta \times T(f)$ is a frequency-dependent real parameter chosen as equal to 0 in the bandwidth and as having a large value β outside of it, with a continuous transition between the two domains. In the following, a weight $\beta = \int_0^{\frac{f_s}{2}} |Y(f)|^2 df$, which corresponds to the energy of the signal to be inverted, has been chosen. In practice, transitions between the two domains can be simple linear functions or \mathcal{C}^∞ Gevrey functions. An example of such a function defining a \mathcal{C}^∞ transition between $T(f_a) = 0$ and $T(f_b) = 1$ is:

$$\forall f \in]f_a, f_b[\quad T(f) = \frac{1}{2} \left[1 + \tanh \left(\frac{1}{f_a - f} + \frac{1}{f_b - f} \right) \right] \quad (4.16)$$

The application of this procedure leads to $y(t)$ that can be considered as an inverse of $e(t)$ in the sense of convolution in the frequency range $[f_1, f_2]$.

Kernel Identification in the temporal domain

After convolving the output of the cascade of Hammerstein models $s(t)$ given in Eq. (4.12) with $y(t)$, one obtains :

$$y * s(t) = \sum_{i=1}^N d * g_i(t + \Delta t_n) = \sum_{i=1}^N \tilde{g}_i(t + \Delta t_n) \quad (4.17)$$

where $\tilde{g}_i(t)$ corresponds to $g_i(t)$ convolved with $d(t)$, *i.e.* to $g_i(t)$ filtered by a bandpass filter in the frequency band $[f_1, f_2]$. The constant K , present in Eq. (4.12), has thus been filtered by $d(t)$. Moreover, if the system under study has no significant behaviour outside of $[f_1, f_2]$, then $\tilde{g}_i(t) = g_i(t)$.

In Fig. 4.2, $y * s(t)$ is represented. Because $\Delta t_n \propto \ln(n)$ and $f_2 > f_1$, the higher the order of linearity n the more in advance the corresponding $\tilde{g}_n(t)$ will be. Thus, if T_m is long enough, the different $\tilde{g}_n(t)$ will not overlap. They are then easy to separate by windowing in the time domain. The separation of the contribution of the different orders of non-linearity by using exponential sweeps, which is mathematically demonstrated here, is already experimentally well known in the audio community [Craven & Gerzon 1992, Norcross & Vanderkooy 1995, Griesinger 1996].

Next, using Eq. (4.18), the family $\{\tilde{h}_n(t)\}_{n \in [1, N]}$ of the Kernels of the cascade of Hammerstein models describing the behaviour of the system in the frequency band $[f_1, f_2]$ can be fully extracted.

$$\begin{pmatrix} \tilde{h}_1(t) \\ \vdots \\ \tilde{h}_N(t) \end{pmatrix} = \mathbf{A}_c^T \begin{pmatrix} \tilde{g}_1(t) \\ \vdots \\ \tilde{g}_N(t) \end{pmatrix} \quad (4.18)$$

\mathbf{A}_c is the Chebyshev matrix defined in the appendix A without its first column and its first row and $(.)^T$ stands for matrix transposition. The first column and the first row of matrix \mathbf{A} have been removed here as there is no continuous components here.

As $\tilde{h}_n(t) = d * h_n(t)$, if the system under study has no significant behaviour outside of $[f_1, f_2]$, then $\tilde{h}_n(t) = h_n(t)$. In most vibration application, systems are designed for a given frequency band (typically

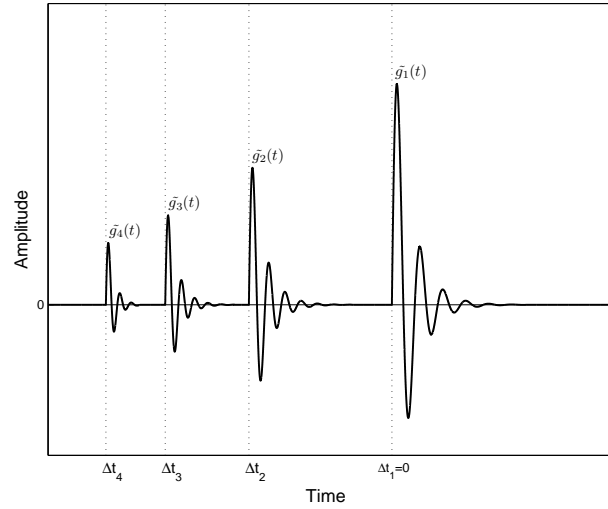


Figure 4.2: Separation of the different orders of non-linearity after convolution with $y(t)$.

[20Hz, 20kHz] for audio applications). The border frequencies f_1 and f_2 can thus generally be selected to identify the real Kernels $h_i(t)$. If it is not possible, Kernels are identified between f_1 and f_2 and thus are only a band-limited version of the real Kernels.

4.2.4 Practical implementation

In this section, the practical discrete-time implementation of the method presented in section 4.2.3 is described.

Overview of the method

In Fig. 4.3, a global overview of the procedure is given. It can be decomposed in the following steps :

1. Design of the input sweep $e(t)$ using Eq. (4.8). The choices of f_1 , f_2 and T are discussed in section 4.2.4.
2. Playing $e(t)$ and recording $s(t)$. The sampling frequency f_s must be chosen to avoid any aliasing effects caused by the digital to analog converter in the frequency range of interest $[f_1, f_2]$.
3. Generation of the inverse filter $y(t)$ according to Eq. (4.15). A convenient way to implement this filter is described in section 4.2.3.
4. Convolution of the output signal $s(t)$ with the inverse filter $y(t)$ as in Eq. (4.17). This can be done in the frequency domain with a sufficient number of points to avoid temporal aliasing.
5. Windowing in the temporal domain to obtain $\{\tilde{g}_k(t + \Delta t_k)\}_{k \in [1, N]}$ (cf. Fig. 4.2). Rectangular windows can be chosen to separate the different orders of non-linearity. N is the highest desired order in the cascade Hammerstein model. Methods to choose N and its influence are shown in sections 4.2.5 and 4.2.5.
6. Temporal shift of the different orders of non-linearity to recover $\{\tilde{g}_k(t)\}_{k \in [1, N]}$. A shift of a non-integer number of samples can be performed with a phase shift in the frequency domain.

7. Multiplication with A_c^T to access $\{\tilde{h}_k(t)\}_{k \in [1, N]}$, according to Eq. (4.18). A_c is the Chebyshev matrix defined in the appendix A without the first column and the first row. The matrix A_c^T of order 8, which is sufficient in practice, is explicitly given in the appendix A.

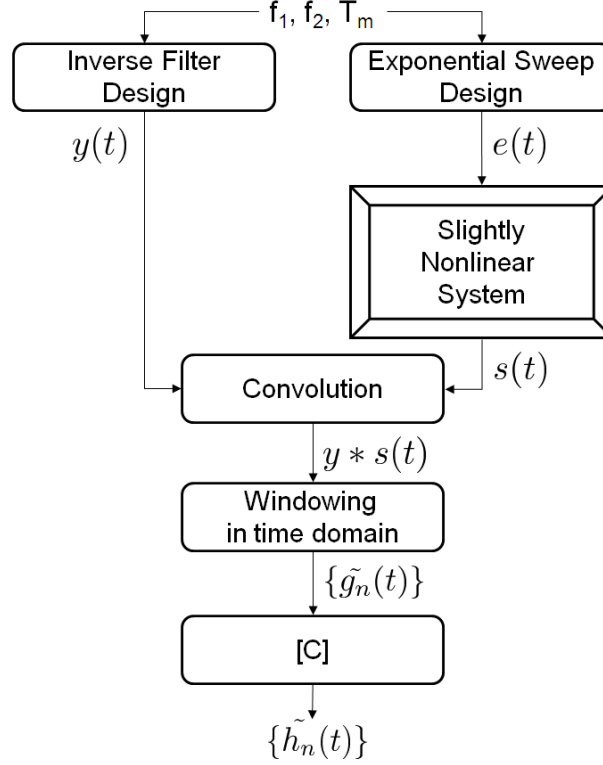


Figure 4.3: Overview of the method used to identify a cascade of Hammerstein models.

Choice of the parameters (f_1 , f_2 , T and N)

For satisfactory measurements, the sweep parameters f_1 , f_2 , T and N must be well chosen. These choices must be made considering the following aspects:

- The frequencies f_1 and f_2 must be chosen such that the interesting behaviour of the system under study is in the frequency range $[f_1, f_2]$.
- The influence of noise on the identification results should also be minimized [Morise et al. 2007]. By itself, the exponential sweep rejects correctly uncorrelated noise in quiet environment [Stan et al. 2002]. Moreover its energy repartition in frequency is often adapted to the ambient noise [Muller & Massarani 2001, Griesinger 1996]. The choice of this signal is thus interesting from this point of view as will be seen in section 4.2.5. If an excellent signal to noise ratio (SNR) is needed, the longer T , the better the SNR after step 4 will be at a given amplitude of the input signal.
- The number of points to be convolved at step 4 is limited by the available computational power. Thus, as T increases, the calculation time will increase. T should not be too large in order to avoid long calculation times.

- N should not be underestimated in order to guarantee good accuracy in identification. The optimal N is reached when it is impossible to extract the corresponding N^{th} impulse response from the background noise. This will be shown in section 4.2.5.
- The different peaks $g_k(t)$ which appear in the temporal domain after the convolution with the inverse signal (step 4, see Fig. 4.2) must not overlap each other. The global decay time of the system τ_{global} is an upper bound of the decay times of each order of non-linearity. Parameters f_1 , f_2 and T such that $\Delta t_N - \Delta t_{N-1} > \tau_{\text{global}}$ will thus avoid overlapping of the different orders of non-linearity [Muller & Massarani 2001, Majdak et al. 2007]. Because $\Delta t_N = \frac{T_m \ln N}{\ln f_2/f_1}$, considering the chosen value for N , T must be chosen to be long enough and f_2/f_1 not so large in order to respect the previous condition.

4.2.5 Validation of the method

In this section, the proposed method of identification is tested on a simulated cascade of Hammerstein models.

Design of the simulated system

A cascade of Hammerstein models of order $N = 4$ has been chosen for simulation purposes. This system is fully represented by its 4 Kernels $h_1(t)$, $h_2(t)$, $h_3(t)$ and $h_4(t)$. For these Kernels, which correspond to linear subsystems, impulse responses of low order ARMA filters (2 poles and 2 zeros) with a roughly 5 ms decay time have been chosen. Parameters of the simulated system are given in Table 4.1.

n	$f_{\text{zeros}}(kHz)$	f_{zeros}/f_s	$ p_{\text{zeros}} $	$f_{\text{poles}}(kHz)$	f_{poles}/f_s	$ p_{\text{poles}} $	Gains
1	0.15	7.81×10^{-4}	0.95	1.5	7.81×10^{-3}	0.95	10^{-1}
2	0.4	2.1×10^{-3}	0.97	2	1.04×10^{-2}	0.96	10^{-2}
3	2	1.04×10^{-2}	0.93	0.1	5.2×10^{-4}	0.97	10^{-3}
4	10	5.21×10^{-2}	0.92	0.5	2.6×10^{-3}	0.95	10^{-5}

Table 4.1: Parameters used for the simulation of the cascade of Hammerstein models of order $N = 4$.

Identification without noise

The method presented in section 4.2.3 and implemented as described in section 4.2.4 has been applied here with the parameters given in Table 4.2 in order to identify the different Kernels of the system.

The magnitude and phase of the frequency responses of the original and estimated Kernels $H_1(f)$, $H_2(f)$, $H_3(f)$ and $H_4(f)$ are shown in Fig 4.4. The estimated Kernels are very close to the original ones over almost the entire frequency range $[f_1, f_2]$. For the frequency regions close to f_1 and f_2 , the estimated Kernels deviates slightly from the original ones, especially the highest orders. This illustrates the limits of the band-limited inverse filter defined by Eq. (4.15).

For a finer analysis, the relative errors in dB, defined in Eq. (4.19), are given for the four estimated Kernels in Fig. 4.5. As magnitude estimation errors and phase estimation errors are included in this relative error, this is a more compact way to access estimation errors.

Parameter	Value	Normalized value
f_1	20 Hz	$f_1/f_s = 1.04 \times 10^{-4}$
f_2	20 kHz	$f_2/f_s = 1.04 \times 10^{-1}$
f_s	192 kHz	
T	15 s	$T \times f_s = 2.88 \times 10^6$ samples
N	4	

Table 4.2: Parameters chosen to identify the simulated system.

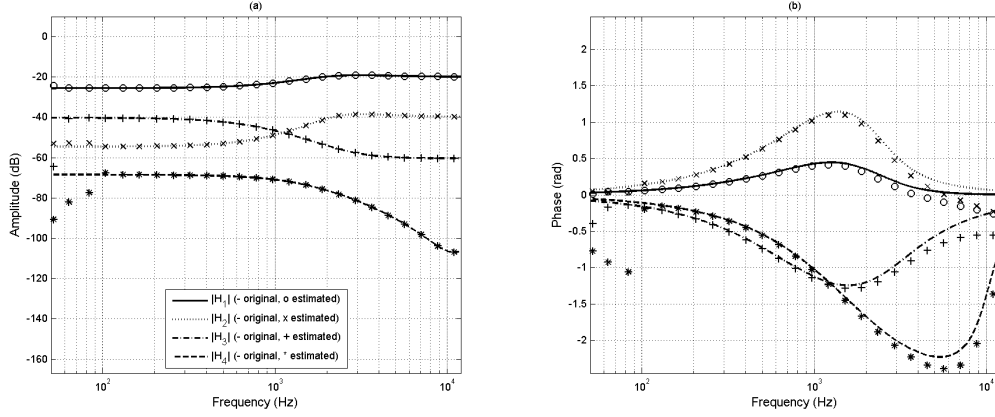


Figure 4.4: Magnitude (a) and phase (b) of the frequency responses of each Kernel of the simulated system. Originals are shown in solid or dotted lines and estimations with o,+,* or x.

$$\gamma_k(f) = 20 \log_{10} \left| \frac{H_k^{\text{orig}}(f) - H_k^{\text{meas}}(f)}{H_k^{\text{orig}}(f)} \right| \quad (4.19)$$

In Fig. 4.5, it can be seen that the relative estimation error is lower than -20 dB over a large portion of the frequency range $[f_1, f_2]$. The consequences of the errors made in magnitude and phase near the border of $[f_1, f_2]$ are clearly visible. To avoid these side effects, the frequency range of the sweep can be chosen larger than the frequency range of interest, depending on the desired accuracy. Near the poles and zeros of the ARMA filters, estimation errors also increase slightly.

The method proposed here gives very good results for identification purposes over a given frequency range, without added noise.

Sensitivity to noise

In a second step, the influence of noise on the estimated Kernels has been studied on the simulated system. A white Gaussian noise at different levels has been added to the output $s(t)$ of the system under study (see Fig. 4.3). Signal to noise ratios relatively to the input (SNR_I) and before convolution (SNR_B), *i.e.* before step 4 (Sec. 4.2.4 and Fig. 4.3) are presented in Table 4.3. Kernel to noise ratios (SNR_A) after step 7 are also given in Table 4.3. SNR is understood here as the ratio between the root-mean-square (RMS) level of the signal (or of the Kernel for SNR_A) in absence of noise and the RMS level of noise in absence

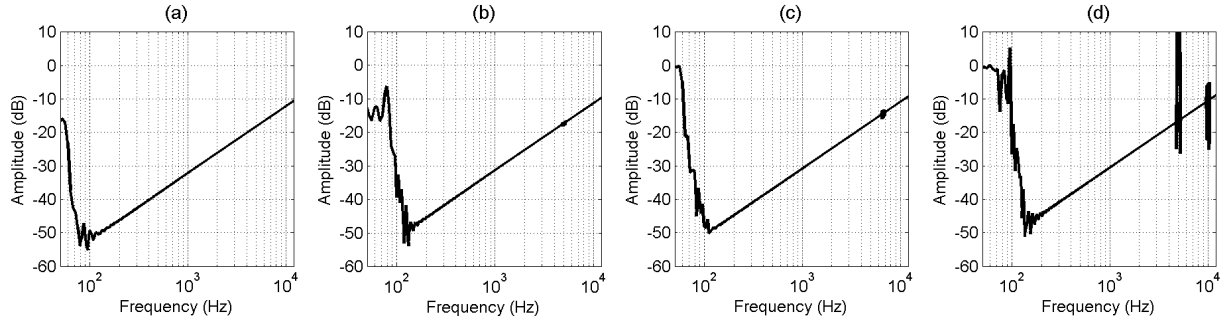


Figure 4.5: Relative errors made during the estimation of the different Kernels. (a) : $H_1(f)$, (b) : $H_2(f)$, (c) : $H_3(f)$, (d) : $H_4(f)$.

of signal (respectively in absence of Kernel). Signals are recorded at the input of the system for SNR_I and at the output of the system for SNR_B . Kernels are taken after the complete identification procedure for SNR_A . SNR_A corresponding to each identified Kernel is given individually. The different Kernels of the system under study have been estimated using the parameters of Table 4.2 in the different noise conditions. In Table 4.3, it can be seen that SNR_A for $N = 1$ is 13.6 dB higher than SNR_B . This confirms the fact that exponential sine sweeps reject a great part of the uncorrelated noise.

SNR _I (dB)	SNR _B (dB)	SNR _A (dB)			
		$N = 1$	$N = 2$	$N = 3$	$N = 4$
37	15.2	29.8	0.4	-7.1	-38.6
57	35.2	49.8	20.4	12.8	-18.6
77	55.2	68.8	40.4	32.8	1.4

Table 4.3: Signal to noise ratio relatively to the input (SNR_I), before convolution (SNR_B) and after the complete identification procedure (SNR_A). SNR_A is indicated relatively to the n^{th} Kernel ($n \in \{1, 2, 3, 4\}$).

In Fig. 4.6, the relative estimation errors, according to Eq. (4.19), are given for the different Kernels and for the different noise levels. It can be seen that the noise level has an influence on the quality of the estimation. For each Kernel, when the noise level is decreasing, the relative estimation error is decreasing too.

For a SNR_I of 37 dB (SNR_B of 15.2 dB), only the estimation of the first Kernel is acceptable. The three other Kernels, which have gains at least 20 dB lower than the first one (see Fig. 4.4), cannot be estimated correctly in this case. Let's consider in Tab. 4.3 the SNR_A for $n \geq 2$. They are all around or below 0 dB. That means that after identification, resulting noises have a RMS level equal to, or higher than, that of identified Kernels. This explains why $H_2(f)$, $H_3(f)$ and $H_4(f)$ cannot be estimated correctly in this case.

When SNR_I is 57 dB (SNR_B of 35.2 dB), the SNR_A corresponding to the second and third Kernels are higher than 10 dB. The second and third Kernels, $H_2(f)$ and $H_3(f)$, are then correctly estimated. But the last Kernel, with $\text{SNR}_A = -18.6$ dB is still not identified.

When the SNR_I is 77 dB (SNR_B of 55.2 dB), the SNR_A corresponding to the fourth Kernel will be 1.4 dB, and the fourth Kernel $H_4(f)$ is then finally correctly estimated.

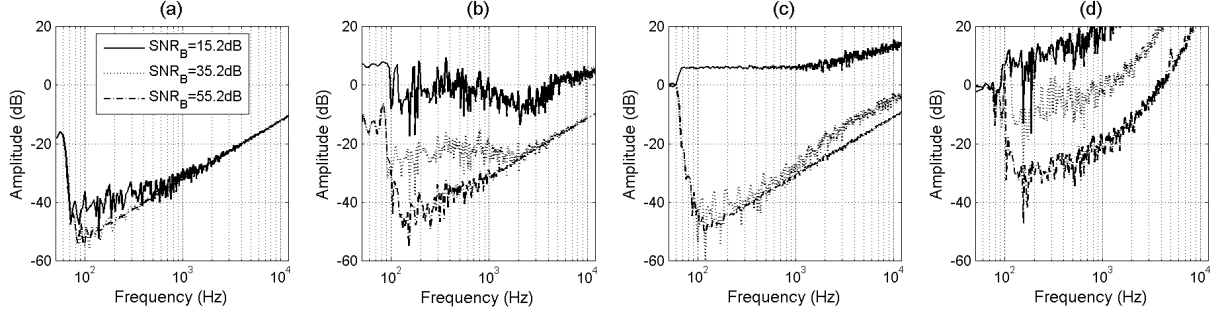


Figure 4.6: Relative errors made during the estimation of the different Kernels for different noise levels. (a) : $H_1(f)$, (b) : $H_2(f)$, (c) : $H_3(f)$, (d) : $H_4(f)$.

So, with the parameters given in Table 4.2 applied to the chosen system, it seems that any Kernel can be correctly estimated until its SNR_A reaches $\simeq 0$ dB. Otherwise, the corresponding Kernel is completely mixed with noise and no information can be extracted. In practice, this defines a measurability criterion (see [Chatterjee & Vyas 2003]) that can be used to estimate the optimal value of N for a given amplitude of the input signal.

Influence of the assumed order of non-linearity

A study of the influence of the order of non-linearity N on estimation has been conducted on the simulated system. The simulated system of order 4 has been identified using the parameters of Table 4.2 but with different assumed orders of non-linearity N ranging from 2 to 6. A white Gaussian noise with a SNR_I of 80 dB (SNR_B of 57.2 dB) has also been added to the output $s(t)$ of the simulated system. In Fig. 4.7, the relative estimation errors made on the Kernels estimated with the different orders of non-linearity are shown. The case $N = 4$ will be the reference as it is the exact order of non-linearity of the simulated system.

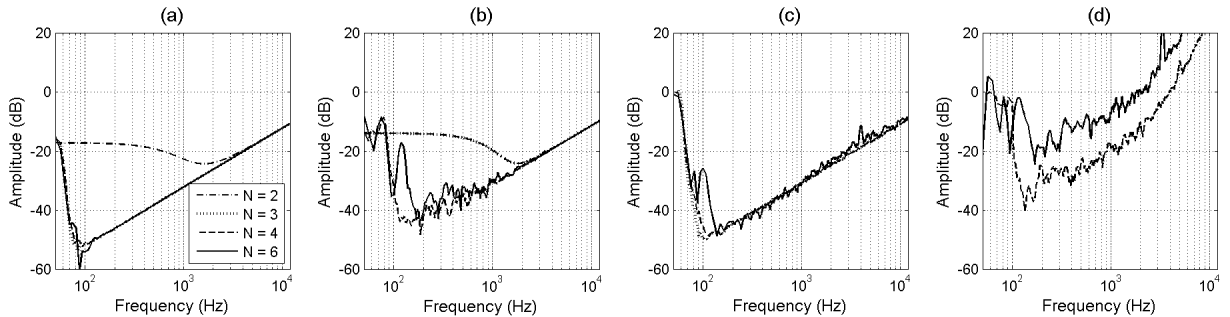


Figure 4.7: Relative errors made during the estimation of the different Kernels for different chosen orders of non-linearity N . (a) : $H_1(f)$ estimated with $N = \{2, 3, 4, 6\}$. (b) : $H_2(f)$ estimated with $N = \{2, 3, 4, 6\}$. (c) : $H_3(f)$ estimated with $N = \{3, 4, 6\}$. (d) : $H_4(f)$ estimated with $N = \{4, 6\}$.

If this order of non-linearity is underestimated (cases $N = 2$ and $N = 3$), the method gives inaccurate

results. This is because of the link which exists between the different estimated Kernels $\{h_n(t)\}_{n \in \{1 \dots N\}}$ and the extracted impulse responses $\{g_n(t)\}_{n \in \{1 \dots N\}}$. This link is the matrix A_c^T , as seen in Eq. (4.18). The first coefficients of the matrix A_c^T are given in Eq. (A.7) in the appendix A. By viewing these coefficients, it's obvious that odd Kernels depend only on odd extracted impulse responses and that the same stands for even Kernels. If an impulse response of order n odd (or even) is not taken into consideration, it will have consequences on all the Kernels of order i odd (or even) lower than n . When $N = 2$, the non-linearities of order 3 and 4 are not taken into consideration and induce estimation errors on the Kernels of order 1 and 2. And when $N = 3$, the non-linearity of order 4 is not taken into consideration and induces estimation errors on the Kernel of order 2 only.

On the other hand, if the order of non-linearity N is over-estimated (case $N = 6$), some portion of noise will be interpreted as extracted impulse responses. As a consequence, estimations of the Kernels are slightly less precise than in the reference case. However, as can be seen in Fig. 4.7, the loss in accuracy is acceptable.

Thus, to ensure an estimation which is as close as possible to reality, it is better to choose the order of non-linearity N as large as possible. The upper limit of N is reached when it is impossible to extract the corresponding impulse response from the background noise. As has been shown in section 4.2.5, this case occurs when the SNR_A corresponding to the N^{th} Kernel reaches a certain level ($\simeq 0$ dB for the chosen system and parameters). In practice, experimental SNR_A can be calculated and used to determine the upper limit of N , as in Sec. 4.2.6.

4.2.6 Modelling acoustical transducers with cascade of Hammerstein models

In this section, acoustical transducers are represented by cascade of Hammerstein models and their Kernels are identified using the method presented in section 4.2.3.

Experimental setup

Experiments have been conducted on two acoustical transducers to identify their Kernels in a cascade of Hammerstein models representation. A standard electrodynamic loudspeaker and a prototype panel-type transducer have been tested. The panel-type transducer consists of a light and stiff plate of sandwich material (40 cm \times 60 cm) on the back of which an exciter has been glued. The plate is suspended by elastics at the middle of two of its side to approximate free boundary conditions. All measurements have been made on axis at one meter from the motor of both transducers with a microphone. Measurements have been done in a semi-anechoic room.

Measured cascade of Hammerstein models Kernels

The Kernels corresponding to both systems have been measured using the previously described experimental setup. As the cascade of Hammerstein models is a non-linear model, its Kernels should be independent of the amplitude of measurement. To assess this, measurements of the Kernels corresponding to both transducers have been done using parameters given in Table 4.4 for 10 different amplitudes. Amplitudes were ranging from 74 to 94 dB in pressure at 1 kHz for the electro-dynamic loudspeaker. This corresponds

to normal and relatively high listening levels. For the panel-based transducer, amplitude where higher, ranging from 90 dB to 110 dB in pressure at 1 kHz.

Parameter	Value	Normalized value
f_1	20 Hz	$f_1/f_s = 1.04 \times 10^{-4}$
f_2	20 kHz	$f_2/f_s = 1.04 \times 10^{-1}$
f_s	192 kHz	
T	15 s	$T \times f_s = 2.88 \times 10^6$ samples
N	5	

Table 4.4: Parameters chosen to identify the real system.

Measurability of each Kernels, using the criterion defined in Sec. 4.2.5, has been studied for the different amplitudes at which Kernels have been identified. As it is not possible to experimentally remove noise from measurements, an experimental Kernel to noise ratio (SNR_{XP}) is defined which corresponds to the ratio between the RMS level of the Kernel with noise and the RMS level of noise in absence of Kernel, after the complete identification procedure. This SNR_{XP} can be computed for each Kernel separately and can be interpreted as SNR_A (see Sec. 4.2.5).

For both transducers, SNR_{XP} corresponding to each Kernel are presented in Fig. 4.8 as a function of the measurement amplitude. As expected, the SNR_{XP} for $N = 1$, *i.e.* for the linear transfer function is linear with the measurement amplitude. One can notice that it is not the case for Kernels of order $n \geq 2$. From Fig. 4.8, it can also be seen that for the lower amplitudes, Kernels of order 3, 4 and 5 have a SNR_{XP} close to 0 dB. They are thus not measurable and pollute slightly the other Kernels. As has been seen in Sec. 4.2.5 a limited pollution is acceptable. Moreover, these Kernels become measurable as the measurement amplitude increases. On the other hand, Kernels of order $n \geq 6$ are always hardly measurable for both transducers in the chosen range of amplitude with the present experimental setup. Consequently, the choice of $N = 5$ in the identification procedures (see Tab. 4.4) seems to be a reasonable compromise between pollution of the identified Kernels by noise and incomplete modelling of the system.

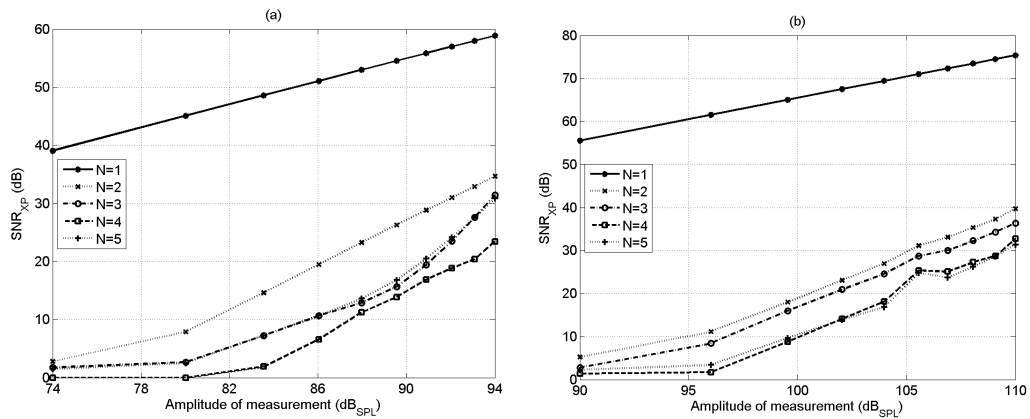


Figure 4.8: Experimental Kernel to noise ratio after the complete identification procedure (SNR_{XP}) as a function of the amplitude of measurement for each Kernel. (a) Electro-dynamic loudspeaker. (b) Panel-based transducer.

Mean measured Kernels and their corresponding standard deviation across measurement amplitude are then given in Fig. 4.9 for the electro-dynamical loudspeaker and in Fig. 4.10 for the panel-based loudspeaker. Due to their different underlying physical principles, the linear and non-linear responses of the two transducers are quite different. The panel-based transducer has a modal behaviour and as a consequence exhibits a linear response with more dips than the electrodynamic loudspeaker. The amplitude of the different Kernels of order $n \geq 2$ decreases with frequency for the electrodynamic loudspeaker, which is consistent with the physical analysis of [Klippel 1996a]. For the panel-based loudspeaker, the amplitude stays globally constant with frequency. The major non-linear phenomena involved in these two transducers do not have the same physical origins.

The variability of the measured Kernels with the excitation amplitude is studied afterwards. The linear part, which is by definition independent of amplitude, exhibits no variations among the different measurements for both transducers. The non-linearities have thus been removed successfully from the linear part using the proposed method. The identified Kernels of order $n \geq 2$ depends slightly on the amplitude at which they have been measured. As a consequence, the assumption that these two transducers can be represented by a cascade of Hammerstein models is a correct approximation in the chosen range of amplitude.

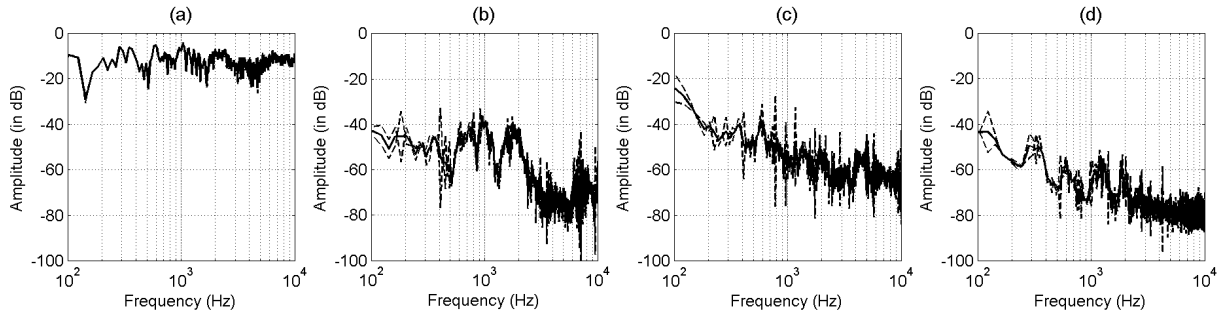


Figure 4.9: Mean measured Kernels of the cascade of Hammerstein models for the electrodynamic loudspeaker (solid line) and the corresponding standard deviations (dashed lines). (a) : $H_1(f)$, (b) : $H_2(f)$, (c) : $H_3(f)$, (d) : $H_4(f)$.

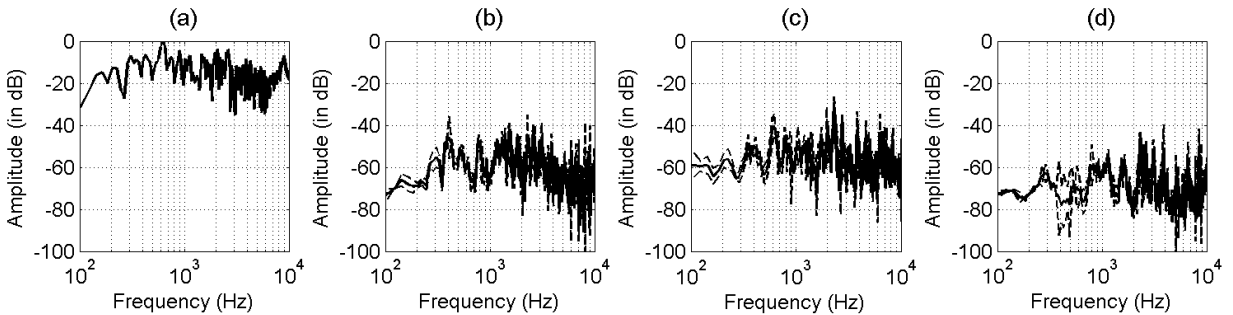


Figure 4.10: Mean measured Kernels of the cascade of Hammerstein models for the panel-based loudspeaker (solid line) and the corresponding standard deviations (dashed lines). (a) : $H_1(f)$, (b) : $H_2(f)$, (c) : $H_3(f)$, (d) : $H_4(f)$.

4.2.7 Prediction of the harmonic distortion generated by transducers

In this section, the previously identified cascade of Hammerstein models will be used to predict the harmonic distortion generated by both transducers.

Link between HD_n , THD and cascade of Hammerstein models

To characterize distortion generated by an acoustic transducer, the following approach is classically adopted. The input of the system is assumed to be sinusoidal and non-linearities generate harmonic components at frequencies higher than the input fundamental frequency. The amplitudes of these harmonics compared to the amplitude of the fundamental are considered as representative of the non-linearity of the transducer. Total harmonic distortion (THD) and harmonic distortion of order n (HD_n) are common tools to quantify this [Czerwinski et al. 2001]. The THD is the square root of the ratio of the power contained in the harmonics to the power contained in the fundamental. The HD_n is the same but for the n^{th} harmonic only.

For a sinusoidal input signal $x(t) = X \cos(2\pi ft)$ which enters a cascade of Hammerstein models identified at the amplitude X_0 , the output signal $z(t)$ can be written as in Eqs. (4.20) by using Eq. (4.2) and (4.7).

$$z(t) = \sum_{n=1}^N |\Gamma_n(X, nf)| \cos[2\pi nft + \angle(\Gamma_n(X, nf))] \quad (4.20a)$$

$$\text{with } \Gamma_n(X, f) = \sum_{k=1}^N \left(\frac{X}{X_0}\right)^k C(k, n) H_k(f) \quad (4.20b)$$

THD and HD_n can thus be directly identified from Eq. (4.20) and expressed as :

$$HD_n(X, f) = \left| \frac{\Gamma_n(X, nf)}{\Gamma_{\text{Tot}}(X, f)} \right| \quad (4.21a)$$

$$THD(X, f) = \sqrt{\sum_{n=2}^N \left[\frac{\Gamma_n(X, nf)}{\Gamma_{\text{Tot}}(X, f)} \right]^2} \quad (4.21b)$$

$$\text{with } \Gamma_{\text{Tot}}(X, f) = \sqrt{\sum_{n=1}^N [\Gamma_n(X, nf)]^2} \quad (4.21c)$$

The knowledge of the Kernels in the frequency range $[f_1, f_2]$ allows for the direct computation of the THD and HD_n using Eqs. (4.21). This can be done for any value of input amplitude X and for any frequency f in $[f_1, f_2]$.

Prediction of HD_n and THD at a given amplitude

Using the different sets of Kernels measured in the previous section, the HD_n and THD for the two transducers has been predicted using Eqs. (4.21). To compare with predictions, traditional measurements

using pure tones have been done using the experimental protocol depicted in section 4.2.6. HD_n and THD have been measured this way for 50 frequencies between 50 Hz and 12 kHz. This has been done for 10 different amplitudes ranging from 74 to 94 dB in pressure for the electro-dynamic loudspeaker and from 90 to 110 dB for the panel-based one.

In Figs. 4.11 and 4.12 the predictions for the total harmonic distortion (THD), and for the harmonic distortion of order 2 and 3 (HD_2 and HD_3) made using equations (4.21) are shown for the electrodynamic loudspeaker and the panel-based loudspeaker. It can be seen that the agreement between measured and predicted data is satisfying over the entire frequency range for the electrodynamic loudspeaker (Fig. 4.11). For the panel-based loudspeakers, the agreement is also good (Fig. 4.12). Below 200 Hz the predictions sometime underestimate HD_2 , HD_3 .

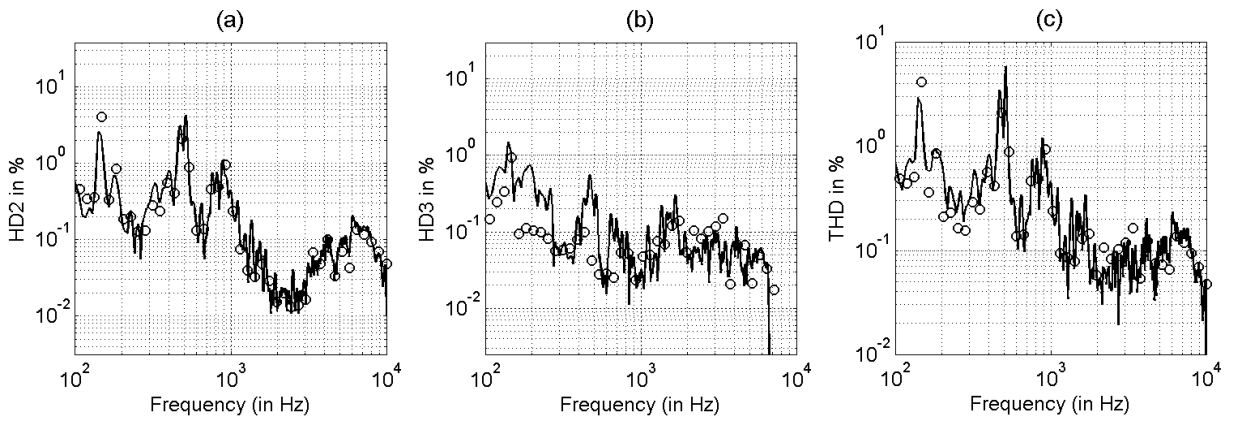


Figure 4.11: Comparisons between measurements (circles) and predictions (solid lines) at 85 dB for the HD_2 , HD_3 and THD of the electrodynamic loudspeaker. Kernels identified at 86 dB have been used for predictions with $N = 5$. (a) Harmonic distortion of order 2, HD_2 . (b) Harmonic distortion of order 3, HD_3 . (c) Total harmonic distortion, THD.

However, evaluation of harmonic distortion using series does not necessary converge to the desired result [Chatterjee & Vyas 2003, 2000]. Indeed, the number of terms N to be used in the series of Eq. (4.21) has to be carefully chosen for predictions to converge to measurements. To study that point, THD has been predicted using 2, 3, 4 or 5 terms in Eqs. (4.21) and compared to measurements. The mean error in frequency between predictions and measurements for different values of N is presented in Fig. 4.13 for both transducers. It has been computed for the following frequency bands: $[45, 180]$ Hz (octave bands 63 Hz and 125 Hz), $[180, 710]$ Hz (octave bands 250 Hz and 500 Hz), $[710, 2800]$ Hz (octave bands 1 kHz and 2 kHz) and $[2800, 11200]$ Hz (octave bands 4 kHz and 8 kHz). For the electro-dynamic loudspeaker in the two upper frequency bands, as N increases the mean error become lower. For the lower frequency bands, the mean error increases with N until $N = 4$ where it starts decreasing. For the panel-based loudspeaker, mean error globally decreases with N . Thus, the choice of $N = 5$ (see Tab. 4.4) in the identification procedure and for the predictions leads to globally convergent results in the chosen ranges of amplitude and frequency.

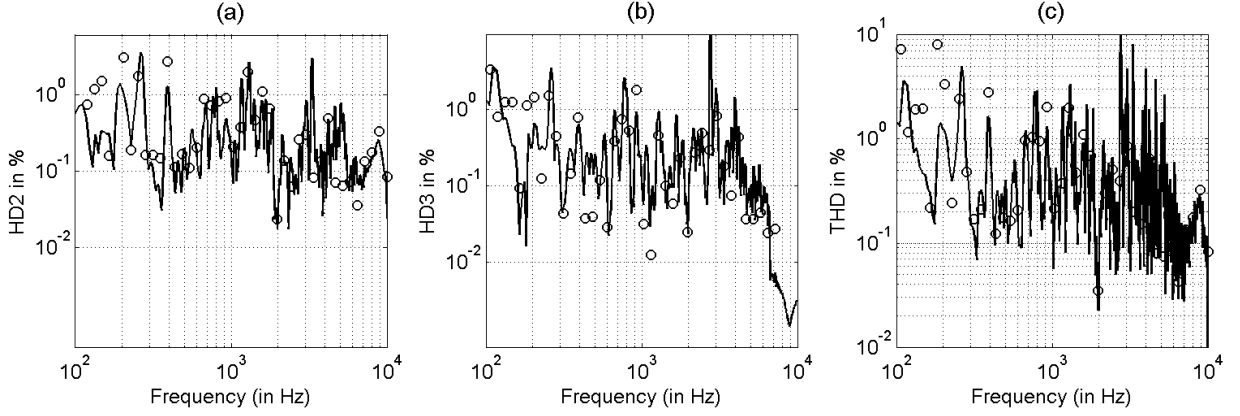


Figure 4.12: Comparisons between measurements (circles) and predictions (solid lines) at 101 dB for the HD_2 , HD_3 and THD of the panel-based transducer. Kernels identified at 105.5 dB have been used for predictions with $N = 5$. (a) Harmonic distortion of order 2, HD_2 . (b) Harmonic distortion of order 3, HD_3 . (c) Total harmonic distortion, THD.

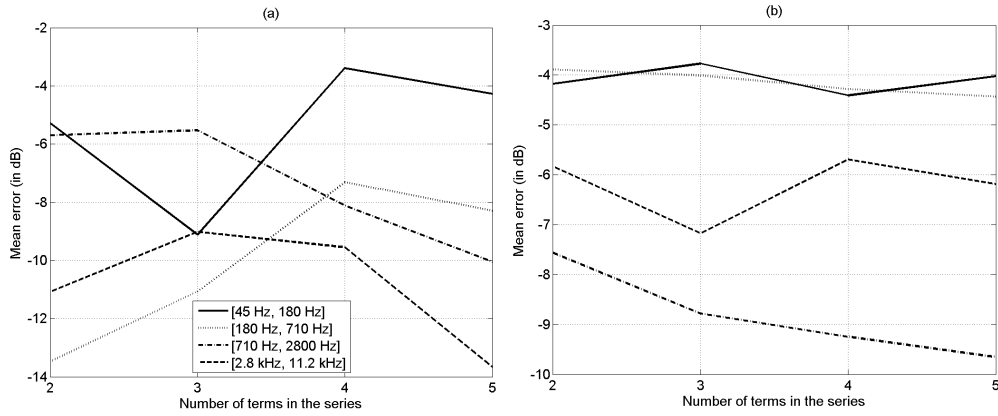


Figure 4.13: Convergence of the prediction error averaged over different frequency bands for the electro-dynamic and panel-based transducers. (a) Electro-dynamic loudspeaker: Kernels identified at 88 dB have been used for predictions at 85.1 dB. (b) Panel-based loudspeaker: Kernels identified at 99.5 dB have been used for predictions at 92.2 dB.

Prediction of HD_n and THD for different amplitudes

To have an overview of the quality of the predictions depending on the amplitude at which Kernels have been measured (X_m) and on the amplitude at which prediction are made (X_p), a mean error has been introduced. This error is defined in Eq. 4.22. This error has been computed for the following frequency bands: $[45, 355]$ Hz (octave bands 63 Hz to 250 Hz), $[355, 2800]$ Hz (octave bands 500 Hz to 2 kHz) and $[2800, 11200]$ Hz (octave bands 4 kHz to 8 kHz). The error in dB in each of these frequency bands is shown for the two transducers in Figs. 4.14 and 4.15.

$$\Delta_{[f_A, f_B]}(X_m, X_p) = \dots \quad (4.22)$$

$$20 \log_{10} \text{mean}_{[f_A, f_B]} \left| \frac{\text{THD}_{\text{meas}}(X_p, f) - \text{THD}_{\text{pred}}(X_m, X_p, f)}{\text{THD}_{\text{meas}}(X_p, f)} \right|$$

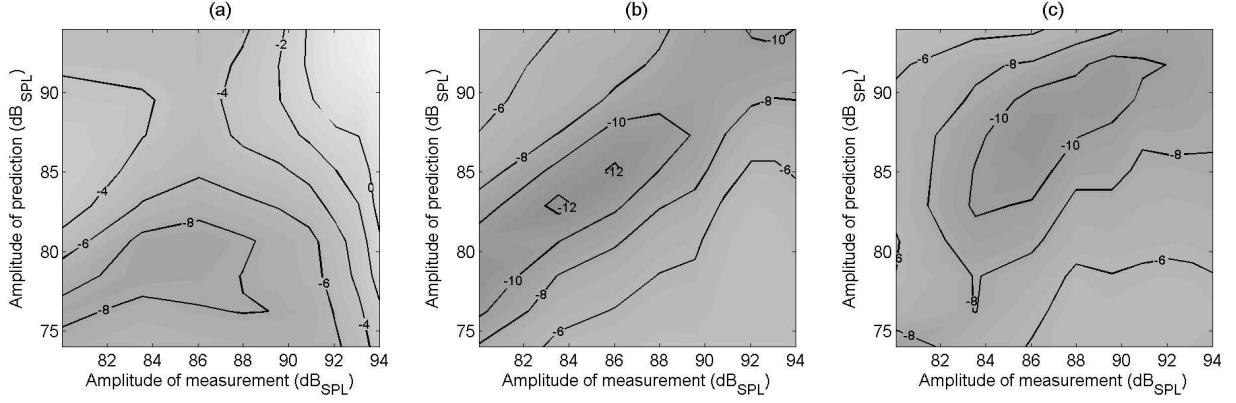


Figure 4.14: Mean error in the frequency band [45, 355] Hz (a), [355, 2800] Hz (b) and [2800, 11200] Hz (c) for the electrodynamic loudspeaker. Amplitude of measurements of the Kernels are given on the x-axis.

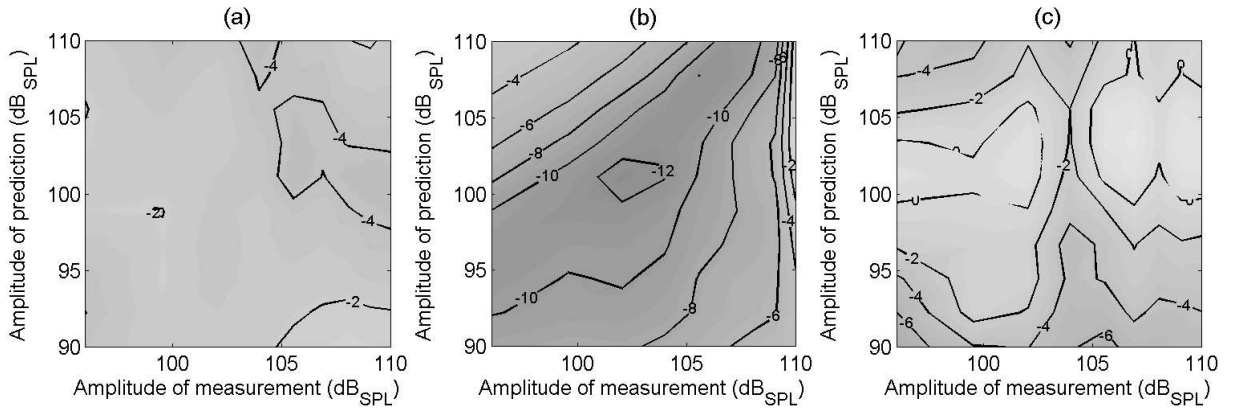


Figure 4.15: Mean error in the frequency band [45, 355] (a), [355, 2800] (b) and [2800, 11200] Hz (c) for the panel-based transducer. Amplitude of measurements of the Kernels are given on the x-axis.

Figs. 4.14 (a), (b), and (c) give the resulting error for the electrodynamic loudspeaker. In Fig. 4.14 (a), the error in the frequency band [45, 355] Hz is shown. It can be seen that this error is acceptable, around -6 dB. The minimum of -8 dB is reached when the THD is predicted for low values of X_p . As the amplitude of prediction X_p increases, the error increases too in this frequency band. In Figs. 4.14 (b) and (c), errors for frequency bands [355, 2800] and [2800, 11200] Hz are shown. Error values in these frequency bands are significantly lower than in the previous one. The minimums of these errors, which are -12 dB and -10 dB, can be seen around the diagonals. Predictions are then precise in these frequency bands.

Figs. 4.15 (a), (b), and (c) give the same errors for the panel-based transducer. In Figs. 4.15 (a) and (c),

the errors in frequency bands $[45, 355]$ Hz and $[2800, 11200]$ Hz are shown. These results are acceptable though less satisfying than the results obtained for the electrodynamic loudspeaker. In Fig. 4.15 (b), errors for frequency band $[355, 2800]$ Hz are shown. Error values in these frequency bands for the panel-type loudspeakers are of the same magnitude as for the electrodynamic loudspeaker, remaining low, around -10 dB.

4.2.8 Discussion

The presented method has been applied to two real acoustic transducers, a classic electro-dynamical loudspeaker and a panel-based transducer. As a complete model describing the non-linear behaviour of these transducers is accessible using the previous method (see section 4.2.6), an analysis of the two systems can be performed from this point of view. The first difference between them appears when viewing their linear response. The panel-based transducer exhibits a well known diffuse behaviour in the high frequency range, whereas the electrodynamic loudspeaker does not. This diffuse behaviour is also present in the Kernels of higher order. For the two transducers, the amplitudes of the higher order Kernels can also be compared. Kernels of the panel-based loudspeaker generally have a lower amplitude than Kernels of the electro-dynamic loudspeaker. For the electro-dynamic loudspeaker, the amplitude of Kernels of order $n \geq 2$ have a tendency to decrease with the frequency, which is coherent with [Kaizer 1987, Klippel 1996a], whereas for the panel-based loudspeaker there is no global variation with frequency. Panel-based transducers thus seems to generate less distortion than electrodynamic loudspeakers, and such distortion is almost constant with frequency.

In section 4.2.7, prediction of the harmonic distortion of order n (HD_n) and the total harmonic distortion (THD) from the identified Kernels have been performed. The originality of the present approach is that an analytical formula involving the identified Kernels and the amplitude at which they have been identified allows one to predict HD_n and THD at different amplitudes. This is an advancement compared to current literature where HD_n and THD are usually predicted only for a given amplitude [Muller & Massarani 2001, Farina 2000, Kaizer 1987]. The results obtained for the two systems when comparing this approach to the traditional one at different amplitudes for HD_n and THD are satisfying. These results validate that cascade of Hammerstein models are a well adapted model.

4.2.9 Conclusion

In this paper a simple and rapid model based procedure to measure non-linearities of a vibrating system has been presented mathematically, validated by simulation and finally applied to two acoustical transducers. Cascade of Hammerstein models have been chosen here to model the non-linearities. It has been shown on simulations that the identification method is very accurate. Harmonic distortion generated by these devices is afterwards precisely predicted using this model.

This method, coming from the audio community, can be of great interest in the more general field of vibrations. In modal analysis, for example, a common limitation to access high frequencies is the signal to noise ratio (SNR). Continuous acoustical excitations with high levels are thus of great interest to increase the SNR, but only if the resulting signals are not polluted by non-linearities. Using the presented method, structures can be acoustically excited at high amplitude levels with the assurance that the non-linear part of

the excitation present in the measurements can be completely removed. This method can thus help to solve practical problems which are commonly encountered in experimentations involving vibrations.

This method can also be interesting for transducer quality assessment. It is now known that traditional non-linear measurements tools (such as HD_n and THD) correlate poorly with subjective experiences [Voishvillo 2006]. However, the present approach gives a fine, input-independent representation of the linear and non-linear characteristics of real transducers. As a consequence, simulations of the non-linear responses of identified or calculated transducers can be easily performed. This can be useful when searching for new criteria to assess the decrease of quality caused by non-linearities in acoustical transducers.

Acknowledgements

The authors would like to thank Thomas Hélié for his helpful advice on this work.

4.3 Discussion

4.3.1 Effets de différents paramètres sur la qualité de l'estimation des noyaux

Dans cette section, les effets de différents paramètres (la durée T du sinus exponentiel, la quantification de T , la méthode de génération du filtre inverse) sur la qualité des noyaux de Hammerstein estimés par la méthode décrite à la section 4.2 sont discutés. Dans les trois sous-sections qui suivent, le système simulé décrit à la section 4.2.5 et dont les noyaux sont connus a été utilisé. Sauf indications contraires, les paramètres du tableau 4.2 sont utilisés lors des estimations.

Durée T du sinus exponentiel

Dans l'utilisation de la méthode décrite à la section 4.2, l'influence du bruit sur les noyaux de Hammerstein estimés doit être minimisée. Le sinus à variation exponentielle de fréquence rejette naturellement bien le bruit dans des environnements calmes [Stan et al. 2002]. Comme l'a montré la section 4.2.5, le choix de ce type de signal est donc intéressant de ce point de vue. Si un excellent rapport *signal sur bruit* est requis, on a intérêt à choisir des sinus exponentiels longs. Pour illustrer ce point, les noyaux du système simulé ont été estimés en utilisant des sinus exponentiels de différentes durées : 1 s, 3 s et 15 s. Un bruit blanc uniforme a été ajouté en sortie du système simulé de façon à obtenir un *rapport signal sur bruit avant convolution* (SNR_B , voir section 4.2.5) de l'ordre de 20 dB. Les erreurs relatives (voir la définition donnée par l'équation (4.19)) résultant des ces différentes estimations sont données à la figure 4.16. On constate alors bien que, au moins dans le cas présent, l'allongement de la durée du sinus exponentiel permet d'obtenir des estimations de meilleure qualité, particulièrement pour les fréquences inférieures à 1 kHz.

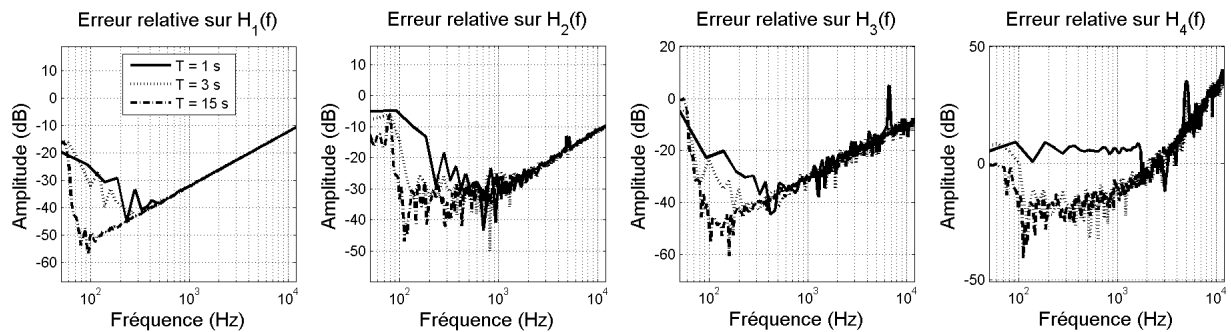


FIG. 4.16 – Erreurs relatives réalisées lors de l'estimation des noyaux de Hammerstein pour différentes valeurs de la durée T du sinus exponentiel.

Quantification de T

Pour que la méthode décrite à la section 4.2 donne des résultats d'estimation de bonne qualité, la longueur T du sinus exponentiel doit impérativement être choisie parmi un jeu de valeurs données analytiquement par l'équation (4.23) :

$$T_m = \left(2m\pi - \frac{\pi}{2}\right) \frac{\ln f_2/f_1}{2\pi f_1} \quad \text{avec } m \in \mathbb{N}^* \quad (4.23)$$

Un mauvais choix de T , c'est à dire un choix de T tel que $T \neq T_m$, induit un terme de phase additionnel $\Delta\Phi$ dans la propriété de phase fondamentale. L'équation (4.11) devient alors :

$$\forall k \in \mathbb{N}^*, \quad \cos[k\Phi(t)] = \cos[\Phi(t + \Delta t_k) + (k - 1)\Delta\Phi] \quad (4.24a)$$

$$\Delta t_k = \frac{T \ln k}{\ln f_2/f_1} \quad (4.24b)$$

$$\Delta\Phi = \frac{\pi}{2} + 2\pi \frac{f_1 T}{\ln f_2/f_1} \quad (4.24c)$$

Les équations (4.17) et (4.18) ne sont alors plus valides ce qui a des conséquences directes sur la qualité des estimations des noyaux $\{h_n(t)\}_{n \in \{1 \dots N\}}$. Pour illustrer ce point, les noyaux du système simulé ont été estimés en utilisant des valeurs de T non-optimales. Les valeurs choisies pour T et les valeurs de déphasage $\Delta\Phi$ correspondantes sont données dans le tableau 4.5. Il est assez impressionnant de constater que de petits écarts autour de la valeur optimale de T peuvent induire d'importantes variations de $\Delta\Phi$. Sur la figure 4.17, les erreurs relatives d'estimations réalisées en utilisant les mauvaises valeurs de T sont comparées à celle faite avec la valeur optimale T_m .

T	$\Delta\Phi$
$T_m = 15.11$ s	0
$T_1 = 15.096$ s	$\pi/12$
$T_2 = 15.024$ s	$\pi/2$

TAB. 4.5 – Valeurs de T testées et valeurs de déphasage $\Delta\Phi$ correspondantes.

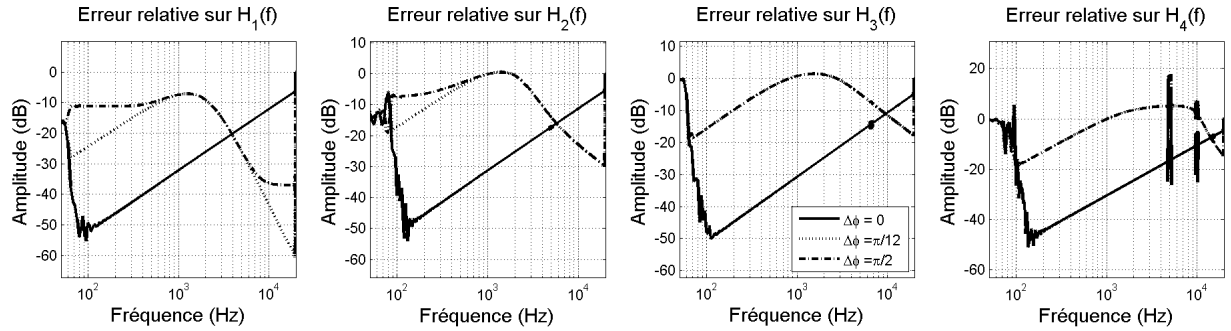


FIG. 4.17 – Erreurs relatives réalisées lors de l'estimation des noyaux de Hammerstein pour différentes valeurs de déphasage $\Delta\phi$.

En ce qui concerne les erreurs relatives réalisées sur $H_3(f)$ et $H_4(f)$, la courbe correspondant à $\Delta\Phi = \pi/12$ est confondue avec celle correspondant à $\Delta\Phi = \pi/2$.

Sur la figure 4.17, le cas $T = T_m$ est considéré comme le cas de référence. Pour les deux mauvaises valeurs de T , la qualité des estimations est globalement moins bonne que pour le cas de référence, spécialement entre 100 Hz et 3 kHz. Pour le noyau d'ordre 1, l'erreur relative peut atteindre $\simeq -10$ dB en choisissant une mauvaise valeur de T . Pour les autres noyaux, les erreurs relatives d'estimation peuvent atteindre $\simeq 0$ dB à certaines fréquences. Le cas où $\Delta\Phi = \pi/2$ est le cas le plus critique. Dans ce cas,

$\cos [3\Phi(t)] = -\cos [\Phi(t + \Delta t_3)]$ au lieu de $\cos [\Phi(t + \Delta t_3)]$. Lors de la reconstruction de $h_1(t)$, la contribution de $g_3(t)$ est ainsi soustraite au lieu d'être additionnée. La baisse de la qualité des estimations pour les autres noyaux et les autres valeurs de $\Delta\Phi$ peut être expliquée par des raisonnements similaires. Cependant, il peut être observé que, curieusement, un mauvais choix de T peut amener à une erreur relative moindre dans les hautes fréquences. Dans ce cas, les erreurs d'estimations induites par le filtre inverse approximé (voir l'équation (4.15)) et les erreurs résultant d'un mauvais choix de T se compensent peut-être mutuellement.

Génération du filtre inverse

Le troisième paramètre à choisir lors de l'utilisation de la méthode décrite à la section 4.2 est la fonction $T(f)$ permettant de générer le filtre inverse $Y(f)$ à partir du signal d'entrée $E(f)$. Cette fonction intervient au travers de $\epsilon(f) = \beta \times T(f)$ dans l'équation (4.25) (voir aussi l'équation (4.15)) :

$$Y(f) \simeq \frac{E^*(f)}{|E(f)|^2 + \epsilon(f)} \quad (4.25)$$

La fonction $T(f)$ vaut 0 dans la bande passante $[f_1, f_2]$ et 1 hors de la bande passante. Une transition discontinue entre ces deux domaines fréquentiels va créer des oscillations dans le domaine temporel, mais une transition lente va limiter la plage fréquentielle sur laquelle les noyaux sont bien estimés. La transition entre ces deux domaines est donc un paramètre à choisir avec précautions et plusieurs possibilités sont disponibles :

- Transition «*Rectangle*» : $T(f)$ est simplement une fonction continue par morceau valant 0 pour $f \in [f_1, f_2]$ et 1 autrement.
- Transition «*Trapèze*» : $T(f)$ est une fonction dérivable par morceau avec des zones de transitions où elle est linéaire au voisinage de f_1 et de f_2 .
- Transition «*Gevrey*» : $T(f)$ est une fonction de classe \mathcal{C}^∞ comme par exemple celle définie par l'équation (4.16).

Pour illustrer les conséquences de ce choix, les noyaux de Hammerstein du système simulé ont été estimés en utilisant des filtres générés avec chacune de ces fonctions $T(f)$. Les erreurs relatives résultant de ces différentes estimations sont données à la figure 4.18. On constate alors que les qualités d'estimations obtenues pour chacune des méthodes sont globalement assez similaires. L'utilisation d'une transition «*Gevrey*», plus *lisse* dans le domaine fréquentiel réduit un petit peu le domaine fréquentiel où l'estimation est de bonne qualité, en particulier en haute fréquences. L'utilisation d'une transition «*Rectangle*», plus *abrupte*, est plus sensible aux résonances du système et donne des résultats de moins bonne qualité aux alentours des pôles et des zéros du système simulé.

4.3.2 Comparaison avec les travaux d'A. Novák *et al.*

En parallèle des travaux décrits dans la section 4.2, une étude similaire à celle présentée ici a été réalisée par A. Novák et ses encadrants au Laboratoire d'Acoustique de l'Université du Maine (LAUM) durant sa thèse [Novák 2009, Novák et al. 2010a,b]. Ces travaux proposent eux aussi une méthode permettant l'estimation des noyaux de Hammerstein d'un système faiblement non-linéaire en utilisant un sweep exponentiel.

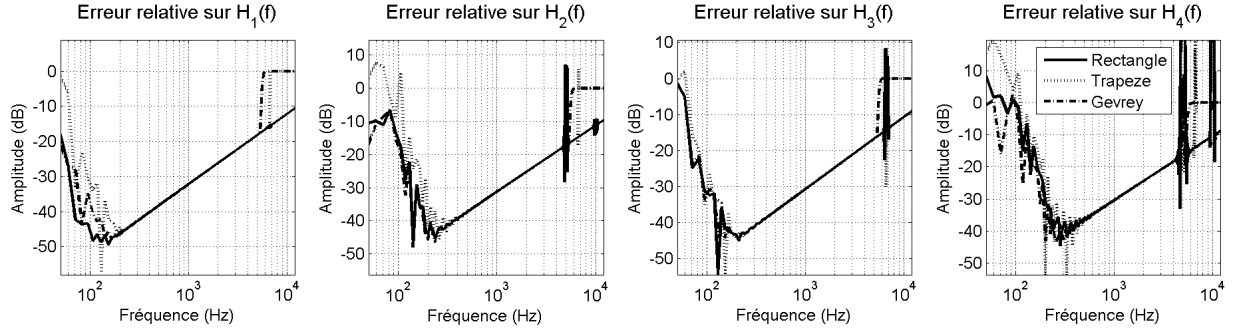


FIG. 4.18 – Erreurs relatives réalisées lors de l'estimation des noyaux de Hammerstein pour différents filtres inverses.

Cependant, même si l'idée directrice guidant Novák *et al.* est identique à celle suivie dans la section 4.2, en pratique, les méthodes développées présentent des différences. Le champ d'application est lui aussi différent étant donné que Novák *et al.* destinent cette méthode à la modélisation d'effets audio. Dans cette discussion, les signaux d'entrées, relations analytiques, filtres inverses et résultats d'estimation obtenus en utilisant chacune des méthodes seront donc comparés.

Signaux d'entrée

Les méthodes proposées par [Novák 2009, Novák et al. 2010a,b] et par la section 4.2 reposent toutes deux sur une classe spécifique de signaux d'entrée : les sweeps exponentiels. Cependant, la durée T de ces sweeps est un paramètre qui ne peut pas être choisi arbitrairement. Cette durée T doit en effet être choisie de façon à ce que $g_k(t)$ (la contribution harmonique d'ordre k) soit en avance temporelle² sur $g_1(t)$ d'un temps $\Delta t_k = \frac{T \ln k}{\ln(f_2/f_1)}$. Ainsi, pour une fréquence de départ f_1 et une fréquence d'arrivée f_2 , la durée T d'un sweep ne peut être choisie que parmi un jeu discret de valeurs analytiquement calculables. L'avance temporelle Δt_k de chaque ordre et le fait qu'elle soit connue permettent ensuite d'extraire par fenêtrage temporel les contributions harmoniques d'ordre k du signal déconvolué. La construction d'un signal d'entrée permettant d'avoir cette propriété est donc primordiale dans les méthodes développées.

Les signaux d'entrée utilisés par les deux approches sont construits suivant différents arguments :

- D'après [Novák 2009, Eq. (4.13-14)] et [Novák et al. 2010a, Eq. (10)], le signal d'entrée de durée $T = L \log(\frac{f_2}{f_1})$ à utiliser est donné par :

$$e(t) = \sin \left\{ 2\pi f_1 L \left[\exp \left(\frac{t}{L} \right) - 1 \right] \right\} \quad \text{avec} \quad L = \frac{1}{f_1} \text{Round} \left[\frac{T f_1}{\log \left(\frac{f_2}{f_1} \right)} \right] \quad (4.26)$$

Ce signal est construit de façon à ce que $\forall k \geq 0, \quad e(\Delta t_k) = 0$ et $\frac{de}{dt}(\Delta t_k) > 0$.

- Le signal d'entrée utilisé par la méthode proposée en section 4.2 est décrit à la section 4.2.3. Ce signal est construit de façon à respecter la propriété de phase fondamentale donné par l'équation (4.10) et de laquelle découlent les Δt_k .

2. On se référera dans la section 4.2.3 à l'équation (4.12) et la figure 4.2 pour la définition de $g_k(t)$, la contribution harmonique d'ordre k , et pour l'illustration des Δt_k .

Sur la figure 4.19, les signaux d'entrée utilisés par chacune des deux approches sont représentés dans les domaines temporel et fréquentiel. D'un point de vue méthodologique, on constate que dans l'approche proposée par [Novák 2009, Novák et al. 2010a,b] les Δt_k permettent la construction du signal d'entrée $e(t)$ alors que dans l'approche proposée en section 4.2 les Δt_k sont une conséquence de la façon dont est construit $e(t)$. Par ailleurs, pour une durée \hat{T} demandée (ici 0.3 s) les durées T retenues sont différentes ($T = 0.32$ s pour la méthode proposée dans [Novák 2009, Novák et al. 2010a,b] et $T = 0.31$ s pour la méthode proposée dans la section 4.2). Enfin, même si dans le domaines fréquentiel les signaux ont un module très similaire, leurs phases sont différentes et leur différence n'est pas une constante.

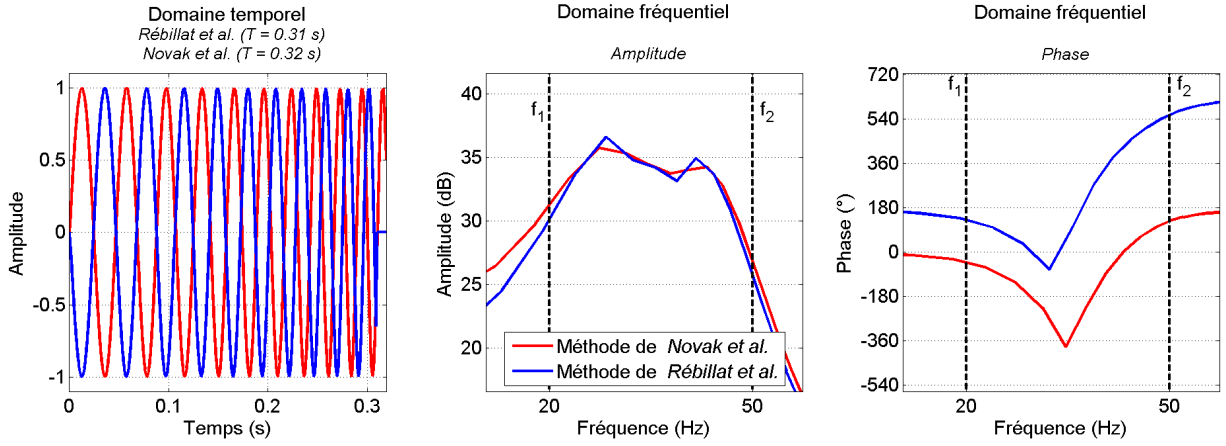


FIG. 4.19 – Signaux d'entrée utilisés dans [Novák 2009, Novák et al. 2010a,b] et dans la méthode proposée en section 4.2 dans les domaines temporel et fréquentiel. Les signaux ont été générés avec $f_1 = 20$ Hz, $f_2 = 50$ Hz, $\hat{T} = 0.3$ s et $f_{\text{éch}} = 1$ kHz.

Relations analytiques entre les composantes harmoniques et les noyaux de Hammerstein

Du fait des différences entre les signaux d'entrée choisis par chacune des méthodes, les matrices \mathbf{L} permettant de faire le lien entre les contributions harmoniques $\{g_n(t)\}_{n \in [1, N]}$ et les noyaux de Hammerstein du système $\{h_n(t)\}_{n \in [1, N]}$ (voir Eq. (4.18) pour la définition de ce lien) diffèrent aussi.

- D'après [Novák 2009, Eqs. (4.50-51)] et [Novák et al. 2010a, Eqs. (48) et (50)], la matrice à construire pour utiliser la méthode proposée par Novák *et al.* est $\mathbf{L}^{\text{Nov}} = (\mathbf{A}^T)^{-1}$ avec :

$$A_{n,m} = \begin{cases} \frac{(-1)^{2n+\frac{1-m}{2}}}{2^{n-1}} \times \binom{n}{\frac{n-m}{2}} & \text{pour } n \geq m \text{ et } (n+m) \text{ pair.} \\ 0 & \text{sinon} \end{cases} \quad (4.27)$$

(4.28)

- La matrice \mathbf{L}^{Reb} utilisée dans la méthode proposée en section 4.2 est construite à partir de polynômes de Chebyshev d'après la procédure explicitée en section A.

À titre d'illustration, les matrices \mathbf{L}^{Nov} et \mathbf{L}^{Reb} générées par chacune des deux méthodes jusqu'à $N = 5$ sont données dans l'équation (4.29). En module, les coefficients de ces matrices sont égaux, mais leurs phases diffèrent. C'est une conséquence des différences de phase entre les signaux d'entrée.

$$\mathbf{L}^{\text{Reb}} = \begin{bmatrix} 1 & 0 & -3 & 0 & 5 \\ 0 & 2 & 0 & -8 & 0 \\ 0 & 0 & 4 & 0 & -20 \\ 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 16 \end{bmatrix} \quad \mathbf{L}^{\text{Nov}} = \begin{bmatrix} 1 & 0 & 3 & 0 & 5 \\ 0 & -2i & 0 & -8i & 0 \\ 0 & 0 & -4 & 0 & -20 \\ 0 & 0 & 0 & 8i & 0 \\ 0 & 0 & 0 & 0 & 16 \end{bmatrix} \quad (4.29)$$

Génération du filtre inverse

Dans les méthodes proposées par [Novák 2009, Novák et al. 2010a,b] et par la section 4.2, la création d'un filtre correspondant à l'inverse du signal d'entrée (au sens de la convolution) est nécessaire pour obtenir les différentes contributions harmoniques d'ordre k (cf. Eq. (4.17)). Ce filtre est important car il va garantir un bon affranchissement vis à vis du signal d'entrée, et donc une bonne estimation des noyaux de Hammerstein du système.

- Dans [Novák 2009, Eq. (4.44)] et [Novák et al. 2010a, Eq. (42)], Novák donne une approximation de ce filtre inverse dans le domaine temporel, sous la forme d'une expression analytique :

$$y(t) = \frac{f_1}{L} \exp\left(-\frac{t}{L}\right) e(-t) \quad (4.30)$$

Cette approximation est construite en utilisant le signal analytique $z_e(t)$ associé à $e(t)$ ne contenant que des fréquences positives. En utilisant ensuite l'approximation de la phase stationnaire, une expression analytique approchée de $Z_e(f)$ peut être dérivée. Il suffit ensuite d'effectuer les mêmes opérations, mais en sens inverse, sur l'expression analytique donnant le spectre $Z_y(f) = \frac{1}{Z_e(f)}$ pour obtenir une approximation du filtre inverse $y(t)$.

- Le filtre inverse $y(t)$ utilisée dans la méthode proposée en section 4.2 est construit en utilisant une méthode numérique de traitement du signal utilisant l'équation (4.15).

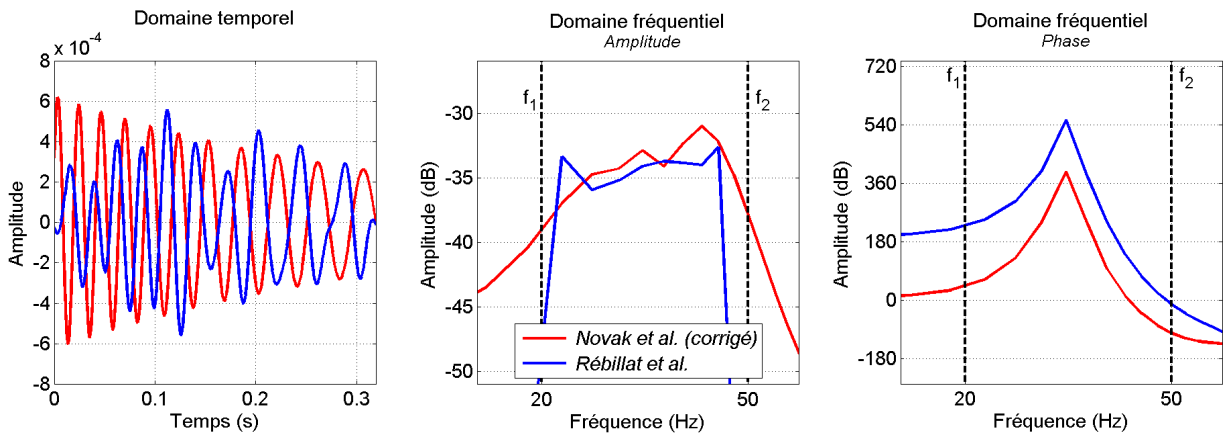


FIG. 4.20 – Filtres inverses utilisés dans [Novák 2009, Novák et al. 2010a,b] et dans la méthode proposée en section 4.2 dans les domaines temporel et fréquentiel. Les filtres ont été générés avec $f_1 = 20$ Hz, $f_2 = 50$ Hz, $T = 0.3$ s et $f_{\text{éch}} = 1$ kHz.

Sur la figure 4.20, les filtres inverses utilisés par chacune des deux approches sont représentés dans les domaines temporel et fréquentiel³. Sur cette figure on constate que même si les deux signaux ont des allures temporelles assez éloignées, ils présentent des allures assez similaires dans le domaine fréquentiel. Une différence de phase notable est encore observée entre les deux filtres ainsi générés.

Comparaison des méthodes sur un système simulé

Au vu des différences existant entre la méthode proposée dans [Novák 2009, Novák et al. 2010a,b] et celle proposée en section 4.2, une comparaison a été effectuée sur un système simulé. Un modèle de Hammerstein en cascade d'ordre $N = 4$ a été choisi pour modéliser le système simulé. Ce système est complètement représenté par ses 4 premiers noyaux de Hammerstein $H_1(f)$, $H_2(f)$, $H_3(f)$, $H_4(f)$. Pour chacun des noyaux de Hammerstein des réponses impulsionnelles de filtres ARMA ayant 2 pôles et 2 zéros ont été choisies. Les paramètres des systèmes simulés sont donnés dans le tableau 4.6.

n	$f_{\text{zeros}}(kHz)$	$ p_{\text{zeros}} $	$f_{\text{poles}}(kHz)$	$ p_{\text{poles}} $	Gains
1	0.15	0.95	1.5	0.97	1
2	0.4	0.97	4	0.95	10^{-1}
3	1	0.93	0.1	0.95	5×10^{-2}
4	2.2	0.92	0.5	0.92	10^{-2}

TAB. 4.6 – Pôles et les zéros choisis pour la simulation des noyaux de Hammerstein du système utilisé pour réaliser la comparaison.

La méthode proposée par Novák *et al.* [Novák 2009, Novák et al. 2010a,b] et celle décrite à la section 4.2 ont ensuite été appliquées à ce système pour estimer les noyaux de Hammerstein du système simulé avec les paramètres donnés dans la Table 4.7. La figure 4.21 représente les amplitudes des noyaux de Hammerstein estimés en utilisant chacune des méthodes.

Paramètre	Valeur
f_1	20 Hz
f_2	10 kHz
f_s	96 kHz
T	10 s
N	4

TAB. 4.7 – Paramètres utilisés pour l'estimation des noyaux de Hammerstein du système simulé en utilisant la méthode proposée en section 4.2 et celle proposée par Novák *et al.* [Novák 2009, Novák et al. 2010a,b].

L'erreur relative moyenne sur l'amplitude commise par chacune des deux méthodes a été calculée par l'Eq. (4.31). Les erreurs relatives moyennes dans la bande de fréquence allant de 30 Hz à 8 kHz commises sur l'amplitude par chacune des deux méthodes et pour chacun des ordres sont données dans la table 4.8.

$$\forall n \in \{1 : 4\} \quad \epsilon_{\text{amp}} = \frac{\text{mean}_{[30 \text{ Hz}, 8 \text{ kHz}]} \left| \frac{|H_n^{\text{est}}| - |H_n^{\text{ori}}|}{|H_n^{\text{ori}}|} \right|}{1} \quad (4.31)$$

3. Un problème de gain observé en utilisant l'équation (4.30) a été compensé en s'assurant que le produit des transformées de Fourier de $e(t)$ et de $y(t)$ était bien égal à 1 dans la bande de fréquence $[f_1, f_2]$.

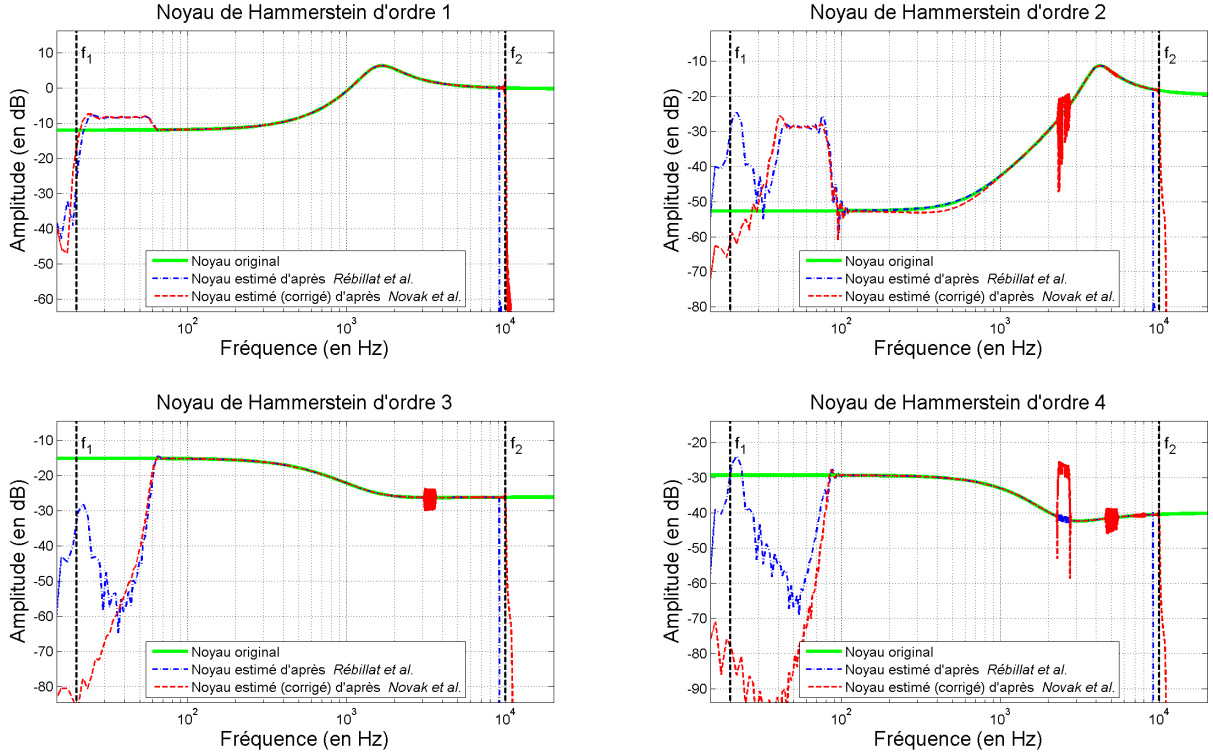


FIG. 4.21 – Estimations obtenues avec chacune des méthodes sur le système simulé.

Méthode	$n = 1$	$n = 2$	$n = 3$	$n = 4$
<i>Rébillat et al.</i>	0.22%	8.8%	0.4%	1%
<i>Novák et al.</i>	0.28%	12.9%	2.19%	25.9%

 TAB. 4.8 – Erreurs relatives moyennes ϵ_{amp} , voir Eq. (4.31), sur l'amplitude commises par chacune des deux méthodes dans la bande de fréquence allant de 30 Hz à 8 kHz et pour chacun des ordres.

Les résultats fournis par la méthode proposée par Novák *et al.* [Novák 2009, Novák et al. 2010a,b] et par celle décrite à la section 4.2 sont de manière générale très proches dans la bande de fréquence $[f_1, f_2]$. Autour de f_1 , les estimations données par les deux méthodes s'éloignent sensiblement du vrai noyau de Hammerstein. La méthode proposée en section 4.2 s'éloigne cependant légèrement moins (voir la table 4.8) du noyau original que celle proposée par Novák *et al.* À proximité de f_2 , les estimations données par la méthode proposée par Novák *et al.* sont plus précises pour ce système que celle donnée par celle décrite à la section 4.2. Aux alentours des pôles et des zéros des noyaux la méthode proposée par Novák *et al.* a tendance à produire des artéfacts que la méthode de la section 4.2 ne produit pas.

Au vu de cette première comparaison il paraît donc difficile de déterminer si une méthode donne de meilleurs résultats que l'autre. Il semble que la méthode proposée par Novák *et al.* [Novák 2009, Novák et al. 2010a,b] permette une estimation des noyaux plus précise en hautes fréquences mais soit plus sensible aux pôles et aux zéros du système à étudier que la méthode proposée à la section 4.2. D'autre part, la méthode proposée à la section 4.2 semble globalement plus précise que celle proposée par Novák *et al.*. Une comparaison plus détaillée de ces deux méthodes (sensibilité au bruit et au choix de N par exemple)

permettrait d'identifier plus clairement les avantages et les inconvénients de chacune d'elle.

4.3.3 Chaîne de systèmes non-linéaires

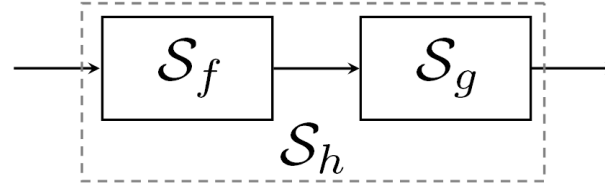


FIG. 4.22 – Chaîne de systèmes non-linéaires.

Il arrive couramment en pratique que l'on cherche à quantifier les non-linéarités introduites par un système S_g dont l'entrée est mesurable, mais pas linéairement contrôlable. L'excitation acoustique (par l'intermédiaire d'un haut-parleur) d'une structure mécanique (une table d'harmonie de piano) dont on veut quantifier la linéarité est un exemple de ce cas de figure.

Comme le montre la figure 4.22, deux systèmes faiblement non-linéaires forment alors une chaîne : un système S_f transmettant le signal émis $x(t)$ jusqu'à l'entrée $y(t)$ d'un autre système S_g . Le système que l'on cherche à étudier est S_g . Ceux que l'on peut mesurer sont S_f et S_h .

Dans cette discussion, la chaîne composée des deux systèmes faiblement non-linéaires S_f et S_g est analysée dans le contexte général des séries de Volterra et dans le contexte plus particulier des modèles de Hammerstein en cascade. La part de non-linéarité apportée par S_g est ensuite estimée, sous certaines hypothèses, en utilisant la méthode d'estimation des noyaux de Hammerstein présentée en section 4.2. Cette méthodologie est enfin appliquée pour quantifier les non-linéarités de la table d'harmonie du piano en collaboration avec K. Ege⁴. Les résultats présentés ici représentent une partie d'un article en préparation pour le *Journal of Sound and Vibration* [Ege et al. 2011].

Séries de Volterra et systèmes de Hammerstein en cascade dans le domaine fréquentiel

Les séries de Volterra permettent d'exprimer de façon exacte la relation existant entre l'entrée $e(t)$ et la sortie $s(t)$ d'un système faiblement non-linéaire [Palm 1978, Boyd & Chua 1985]. Elles peuvent être vues comme une généralisation de la notion de *fonction de transfert*. Dans la continuité de cette notion très utile pour les systèmes linéaires, des *fonctions de transfert non-linéaires* peuvent être définies par transformation de Fourier des noyaux de Volterra :

$$\forall k \in \mathbb{N}^* \quad V_k(f_1, \dots, f_k) = \int_0^{+\infty} \dots \int_0^{+\infty} v_k(\tau_1, \dots, \tau_k) e^{-i2\pi f_1 \tau_1} \dots e^{-i2\pi f_k \tau_k} d\tau_1 \dots d\tau_k \quad (4.32)$$

Les fonctions $\{V_k(f_1, \dots, f_k)\}_{k \in \mathbb{N}^*}$ caractérisent alors le système non-linéaire étudié. L'interprétation des fonctions de transfert non-linéaires reste toutefois très peu intuitive étant donné que la $k^{\text{ième}}$ fonction de transfert $V_k(f_1, \dots, f_k)$ est une fonction de k variables indépendantes.

4. Kerem EGE, Maître de conférences, Laboratoire Vibrations Acoustique, INSA de Lyon, kerem.ege@insa-lyon.fr

Les modèles de Hammerstein en cascade constituent une sous-classe des systèmes modélisables par les séries de Volterra. Si l'on considère un système \mathcal{S}_e modélisable par une cascade de modèles de Hammerstein de noyaux $\{\tilde{E}_k(f)\}_{k \in \mathbb{N}^*}$, alors ses noyaux de Volterra $\{E_k(f_1, \dots, f_k)\}_{k \in \mathbb{N}^*}$ ont la forme suivante (4.33) :

$$\forall (f_1, \dots, f_k) \quad E_k(f_1, \dots, f_k) = \tilde{E}_k(f_1 + \dots + f_k) \quad (4.33)$$

Les noyaux de Hammerstein sont des fonctions d'une seule variable fréquentielle. Leur interprétation est plus simple que celle des noyaux de Volterra, mais les non-linéarités qu'ils permettent de représenter sont aussi moins complexes. Les approximations *physiques* faites en modélisant un système par une cascade de modèles de Hammerstein plutôt que par une série de Volterra ne sont cependant pas encore suffisamment bien comprises et constituent un point qui mériterait une attention spécifique.

Système faiblement non-linéaire équivalent à la chaîne

Soit \mathcal{S}_h le système non-linéaire équivalent à la mise en série des systèmes \mathcal{S}_f et \mathcal{S}_g . Les familles de noyaux de Volterra $\{F_k(f_1, \dots, f_k)\}_{k \in \mathbb{N}^*}$, $\{G_k(f_1, \dots, f_k)\}_{k \in \mathbb{N}^*}$ et $\{H_k(f_1, \dots, f_k)\}_{k \in \mathbb{N}^*}$ décrivent alors complètement les systèmes \mathcal{S}_f , \mathcal{S}_g et \mathcal{S}_h .

Dans ce cas, les noyaux de Volterra de \mathcal{S}_h peuvent s'exprimer analytiquement en fonction des noyaux de Volterra de \mathcal{S}_f et \mathcal{S}_g [Hasler 1999] :

$$\begin{aligned} \forall k \in \mathbb{N}^* \quad H_k(f_1, \dots, f_k) &= \sum_{p=1}^k \sum_{\mathcal{M}_p^k} F_{m_1}(f_1, \dots, f_{m_1}) \times \dots \times F_{m_p}(f_{m_1+\dots+m_{p-1}+1}, \dots, f_k) \dots \\ &\quad \dots \times G_p(f_1 + \dots + f_{m_1}, \dots, f_{m_1+\dots+m_{p-1}+1} + f_n) \\ &\quad \text{avec } \mathcal{M}_p^k = \begin{cases} m_1 + \dots + m_p = k \\ m_1, \dots, m_p \geq 1 \end{cases} \end{aligned} \quad (4.34)$$

Pour les deux premiers termes, l'équation (4.34) se réduit à :

$$\begin{cases} H_1(f_1) = F_1(f_1)G_1(f_1) & (4.35a) \\ H_2(f_1, f_2) = F_2(f_1, f_2)G_1(f_1 + f_2) + F_1(f_1)F_1(f_2)G_2(f_1, f_2) & (4.35b) \end{cases}$$

L'expression (4.35a) démontre alors rigoureusement le résultat intuitif que la *fonction de transfert linéaire* décrivant le comportement d'un système faiblement non-linéaire \mathcal{S}_g suivant un autre système faiblement non-linéaire \mathcal{S}_f est simplement le produit des *fonctions de transfert linéaires* de ces deux systèmes. L'expression (4.35b) permet de déterminer facilement le noyau de Volterra $H_2(f_1, f_2)$ du système \mathcal{S}_h connaissant les noyaux de Volterra des systèmes \mathcal{S}_f et \mathcal{S}_g .

Si \mathcal{S}_f et \mathcal{S}_g sont supposés être modélisables par des cascades de modèles de Hammerstein, alors leurs noyaux de Volterra s'expriment en fonction de leurs noyaux de Hammerstein $\{\tilde{F}_k(f)\}_{k \in \mathbb{N}^*}$ et $\{\tilde{G}_k(f)\}_{k \in \mathbb{N}^*}$ en utilisant la relation (4.33). D'après (4.35), les noyaux de Volterra de \mathcal{S}_h sont alors déduits par :

$$\begin{cases} H_1(f_1) = \tilde{F}_1(f_1)\tilde{G}(f_1) & (4.36a) \\ H_2(f_1, f_2) = \tilde{F}_2(f_1 + f_2)\tilde{G}_1(f_1 + f_2) + \tilde{F}_1(f_1)\tilde{F}_1(f_2)\tilde{G}_2(f_1 + f_2) & (4.36b) \end{cases}$$

Pour que le système \mathcal{S}_h soit modélisable par une cascade de modèles de Hammerstein, il est nécessaire que chacun de ses noyaux de Volterra soit une fonction d'une seule variable, comme stipulé par l'équation (4.33). L'analyse de l'équation (4.36b) révèle que ce n'est pas le cas à cause du terme $\tilde{F}_1(f_1)\tilde{F}_1(f_2)$. Une chaîne de modèles de Hammerstein en cascade ne peut donc pas être représentée par un modèle de Hammerstein en cascade équivalent. Cela n'a donc pas de sens de chercher à estimer les noyaux de Hammerstein de \mathcal{S}_g à partir de ceux de \mathcal{S}_f et de \mathcal{S}_h . Il est cependant possible de chercher à quantifier l'apport non-linéaire du système \mathcal{S}_g dans la chaîne.

Quantification l'apport non-linéaire du système \mathcal{S}_g

Aucune hypothèse n'est à présent faite sur \mathcal{S}_g . \mathcal{S}_f et \mathcal{S}_h sont quand à eux supposés modélisables par des systèmes de Hammerstein en cascade. D'après l'équation (4.36b), il est nécessaire que :

$$\forall f_1, f_2 \in \mathbb{R}^2 \quad |\tilde{F}_1(f_1)\tilde{F}_1(f_2)\tilde{G}_2(f_1 + f_2)| \ll |\tilde{F}_2(f_1 + f_2)\tilde{G}_1(f_1 + f_2)| \quad (4.37)$$

pour que \mathcal{S}_h satisfasse à cette hypothèse.

D'après la figure 4.22, les relations "entrée/sortie" pour les différents systèmes peuvent alors être écrites de la façon suivante :

$$\mathcal{S}_f : \quad Y(f) = \tilde{F}_1(f)X(f) + Y_{NL}^f(f) \quad (4.38a)$$

$$\mathcal{S}_g : \quad Z(f) = \tilde{G}_1(f)Y(f) + Z_{NL}^g(f) \quad (4.38b)$$

$$\mathcal{S}_h : \quad Z(f) = \tilde{H}_1(f)X(f) + Z_{NL}^h(f) \quad (4.38c)$$

D'après les hypothèses faites, il est possible d'appliquer deux fois la méthode décrite précédemment pour estimer les noyaux de Hammerstein des systèmes accessibles, *i.e.* \mathcal{S}_f et \mathcal{S}_h . Les fonctions de transfert linéaires $\tilde{F}_1(f)$ et $\tilde{H}_1(f)$ sont alors connues. De même, des signaux $Y(f)$ et $Z(f)$, contenant des contributions linéaires $\tilde{F}_1(f)X(f)$ et $\tilde{H}_1(f)X(f)$ et non-linéaires $Y_{NL}^f(f)$ et $Z_{NL}^h(f)$ peuvent être calculés.

Quantifier l'apport non-linéaire du système \mathcal{S}_g dans la chaîne revient alors à estimer le rapport de la partie non-linéaire $Z_{NL}^g(f)$ et de la partie linéaire $\tilde{G}_1(f)Y(f)$ en sortie de \mathcal{S}_g à partir de ce qui est connu. En utilisant les relations (4.38), la contribution non-linéaire $NL_{\mathcal{S}_g}^{\text{reel}}(f)$ de \mathcal{S}_g dans la chaîne peut alors être définie :

$$NL_{\mathcal{S}_g}^{\text{reel}}(f) = \frac{Z_{NL}^g(f)}{\tilde{G}_1(f)Y(f)} = \frac{Z(f) - \tilde{G}_1(f)Y(f)}{\tilde{G}_1(f)Y(f)} \quad (4.39)$$

Sous réserve que la chaîne complète \mathcal{S}_h se comporte comme une cascade de modèles de Hammerstein, $NL_{\mathcal{S}_g}^{\text{reel}}(f)$ peut être estimée en utilisant l'équation (4.35a) :

$$\text{NL}_{\mathcal{S}_g}^{\text{esti}}(f) = \frac{\tilde{F}_1(f)Z(f) - \tilde{H}_1(f)Y(f)}{\tilde{H}_1(f)Y(f)} \quad (4.40)$$

Application à une chaîne de systèmes simulée

Dans le but de vérifier la faisabilité d'une telle démarche, une chaîne \mathcal{S}_h composée de deux systèmes non-linéaires \mathcal{S}_f et \mathcal{S}_g a été simulée. La contribution non-linéaire de \mathcal{S}_g à la chaîne a ensuite été estimée en utilisant la démarche exposée à la section 4.3.3 et comparée à sa valeur réelle qui est connue ici.

Des modèles de Hammerstein en cascade d'ordre $N = 4$ ont été choisis pour modéliser les systèmes \mathcal{S}_f et \mathcal{S}_g . Ces systèmes sont donc complètement représentés par leurs 4 premiers noyaux de Hammerstein $\tilde{F}_1(f)$, $\tilde{F}_2(f)$, $\tilde{F}_3(f)$, $\tilde{F}_4(f)$ et $\tilde{G}_1(f)$, $\tilde{G}_2(f)$, $\tilde{G}_3(f)$, $\tilde{G}_4(f)$. Pour chacun des noyaux de Hammerstein des réponses impulsionnelles de filtres ARMA ayant 2 pôles et 2 zéros ont été choisies. Les paramètres des systèmes simulés sont donnés dans le tableau 4.9. Le système \mathcal{S}_h est simplement la mise en série de \mathcal{S}_f et \mathcal{S}_g .

Système	n	$f_{\text{zeros}}(kHz)$	$ p_{\text{zeros}} $	$f_{\text{poles}}(kHz)$	$ p_{\text{poles}} $	Gains
\mathcal{S}_f	1	0.15	0.5	1.5	0.6	3
	2	0.4	0.97	2	0.95	3×10^{-2}
	3	2	0.93	0.1	0.95	3×10^{-3}
	4	10	0.92	0.5	0.92	3×10^{-4}
\mathcal{S}_g	1	0.1	0.6	1.2	0.5	1
	2	0.3	0.95	1.8	0.96	10^{-2}
	3	2	0.93	0.12	0.95	10^{-3}
	4	7	0.92	0.5	0.95	10^{-4}

TAB. 4.9 – Pôles et les zéros choisis pour la simulation des noyaux de Hammerstein des systèmes \mathcal{S}_f et \mathcal{S}_g .

La méthode décrite à la section 4.2 a ensuite été appliquée deux fois avec les paramètres donnés dans la Table 4.10 pour estimer les noyaux de Hammerstein des systèmes \mathcal{S}_f et \mathcal{S}_h .

Parameter	Value
f_1	20 Hz
f_2	10 kHz
f_s	96 kHz
T	10 s
N	4

TAB. 4.10 – Paramètres utilisés pour l'estimation des noyaux de Hammerstein des systèmes \mathcal{S}_f et \mathcal{S}_h .

En utilisant l'équation (4.39), la contribution non-linéaire de \mathcal{S}_g à la chaîne $\text{NL}_{\mathcal{S}_g}^{\text{esti}}(f)$ a ensuite été estimée à partir des mesures et comparée à sa valeur réelle $\text{NL}_{\mathcal{S}_g}^{\text{reel}}(f)$. Les résultats obtenus pour la valeur estimée et pour la valeur réelle sont présentés sur la figure 4.23. À titre informatif, les contributions non-linéaires réelles de \mathcal{S}_f : $\text{NL}_{\mathcal{S}_f}^{\text{reel}}(f)$, et de la chaîne complète *i.e.* \mathcal{S}_h : $\text{NL}_{\mathcal{S}_h}^{\text{reel}}(f)$, ont aussi été tracées.

Sur la figure 4.23, on constate que les valeurs réelles et estimées sont globalement très proches l'une de l'autre dans la bande de fréquence $[f_1, f_2]$. Ceci est vrai dans le cas présent à la fois dans les bandes

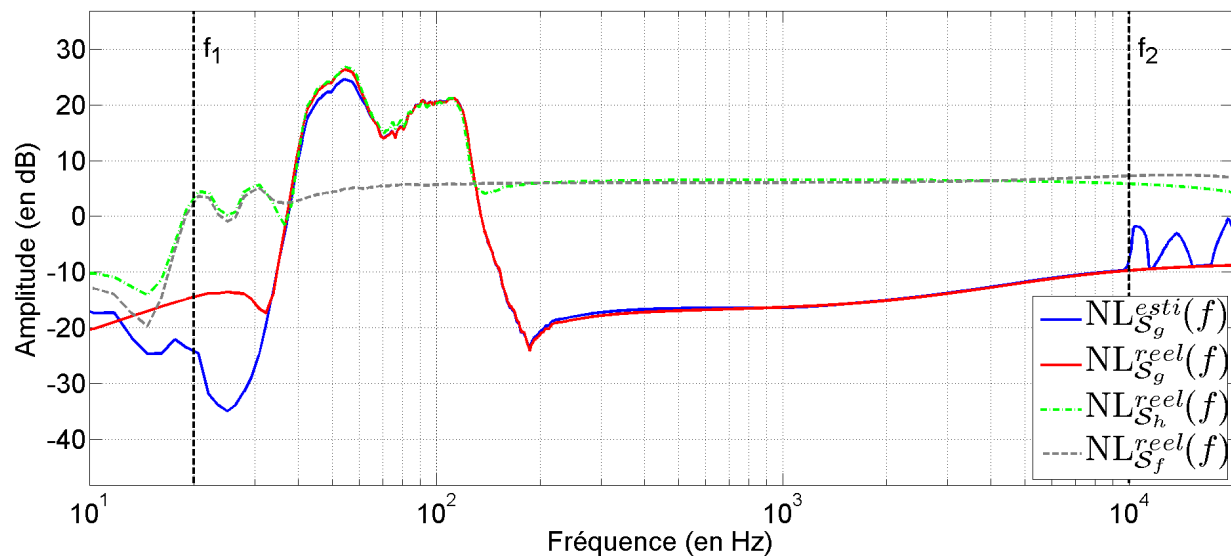


FIG. 4.23 – Comparaison entre les valeurs estimée et réelle de l’apport non-linéaire de S_g à la chaîne de systèmes non-linéaires. Toutes les courbes ont été moyennées avec une fenêtre glissante de largeur logarithmique 0.05.

fréquentielles où le système S_g apporte beaucoup de non-linéarités (*i.e.* de 40 Hz à 100 Hz) ou dans les bandes de fréquences où les non-linéarités sont générées principalement par S_f (*i.e.* de 100 Hz à 20 kHz). L’hypothèse que S_h soit modélisable par une cascade de systèmes de Hammerstein ne semble donc pas être une approximation trop grossière dans l’exemple choisi où les deux systèmes génèrent des non-linéarités de façon analogue. D’autre part on constate aussi que les contributions non-linéaires réelles de S_f et de S_g dépassent quelquefois la contribution non-linéaire totale de S_h , ce qui est théoriquement impossible. Ceci est certainement un artéfact du au moyennage par fenêtre glissante des réponses fréquentielles.

Au vu de ces résultats de simulation, la méthode proposée ici pour étudier l’apport non-linéaire d’un système faiblement non-linéaire inclus dans une chaîne de systèmes faiblement non-linéaires semble donc applicable. De plus amples validations sont cependant nécessaires pour pouvoir conclure de façon plus sûre.

Quantification des non-linéarités de la table d’harmonie du piano

La table d’harmonie du piano est souvent considérée comme linéaire, mais aucune étude n’a à ce jour quantifié à quel degré cette hypothèse était vraie. Pour répondre à cette problématique, la méthode exposée précédemment a été utilisée. Un piano droit dont les cordes ont été amorties a été placé dans une chambre anéchoïque et excité acoustiquement à l’aide d’un haut-parleur électro-dynamique. Dans un premier temps la pression acoustique au voisinage de la table d’harmonie du piano a été mesurée à l’aide d’un microphone et en l’absence du piano. Dans un second temps, l’accélération résultant de l’action de cette pression (supposée uniforme en première approximation) a été mesurée en trois points de la table d’harmonie à l’aide d’accéléromètres. Toutes les mesures présentées dans cette partie ont été réalisées par K. Ege⁵.

5. Kerem EGE, Maître de conférences, Laboratoire Vibrations Acoustique, INSA de Lyon, kerem.ege@insa-lyon.fr

Avec ce dispositif expérimental, la pression reçue par la table d'harmonie est mesurable, mais n'est pas a priori linéairement contrôlable. Le système étudié est donc la mise en cascade de deux systèmes que l'on suppose non-linéaires : le haut-parleur \mathcal{S}_f qui transforme un signal électrique $X(f)$ en un signal de pression $Y(f)$, et la table d'harmonie du piano \mathcal{S}_g qui transforme le signal de pression $Y(f)$ en une accélération $Z(f)$. Il est possible également que les non-linéarités engendrées par le haut-parleur \mathcal{S}_f soient plus importantes que celles générées par la table d'harmonie du piano \mathcal{S}_g .

Le problème peut donc être mis sous la forme proposée par la figure 4.22 et la procédure proposée en section 4.3.3 être appliquée. Les noyaux de Hammerstein des systèmes \mathcal{S}_f et \mathcal{S}_h ont été estimés à l'aide la méthode proposée en section 4.2 avec les paramètres donnés dans la Table 4.11. L'apport non-linéaire $NL_{\mathcal{S}_g}^{esti}(f)$ de la table d'harmonie du piano \mathcal{S}_g peut ensuite être estimée pour chacun des trois accéléromètres à l'aide de l'équation (4.40) .

Parameter	Value
f_1	50 Hz
f_2	4 kHz
f_s	40 kHz
T	26 s
N	4

TAB. 4.11 – Paramètres utilisés pour l'estimation des noyaux de Hammerstein des systèmes \mathcal{S}_f (haut-parleur électrodynamique) et \mathcal{S}_h (haut-parleur électrodynamique et table d'harmonie).

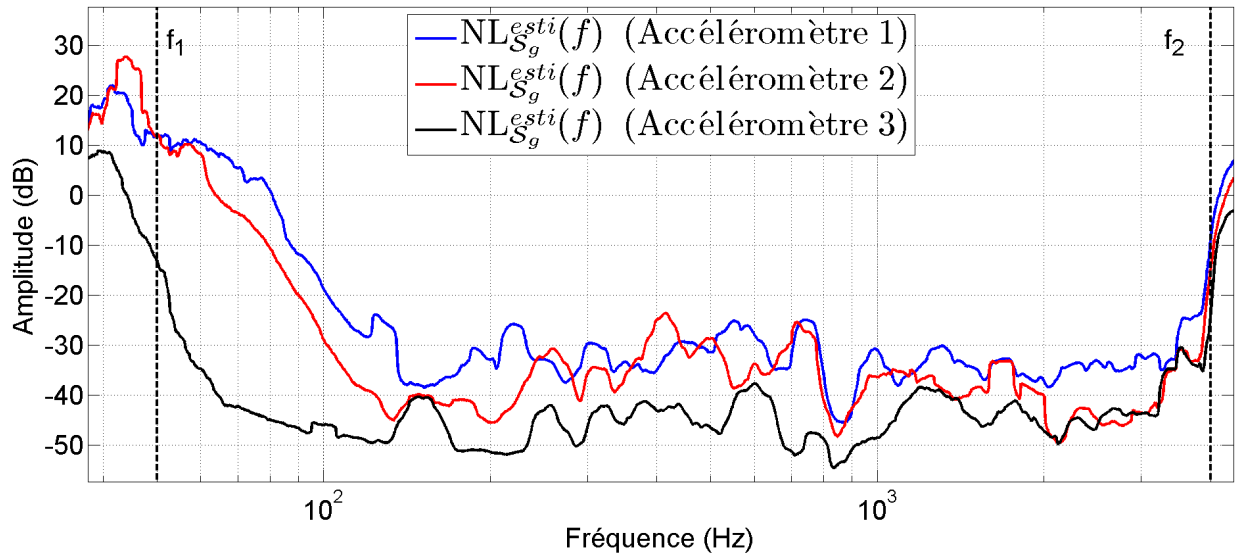


FIG. 4.24 – Valeurs estimées de l'apport non-linéaire de la table d'harmonie du piano \mathcal{S}_g à la chaîne de systèmes non-linéaires. Toutes les courbes ont été moyennées avec une fenêtre glissante de largeur logarithmique 0.05.

Les résultats de l'estimation de l'apport non-linéaire de la table d'harmonie du piano \mathcal{S}_g à la chaîne de systèmes non-linéaires sont présentés à la figure 4.24. On constate dans un premier temps que les trois accéléromètres donnent des valeurs assez similaires. Les différences observées entre les trois courbes pro-

viennent du fait que les accéléromètres sont placés à des endroits différents sur la table d'harmonie et donc subissent des vibrations d'amplitudes différentes. Pour la bande de fréquence [50 Hz, 100 Hz], des valeurs très importantes d'apport non-linéaire sont observées. Or, comme on l'a déjà constaté sur la figure 4.23, les estimations de l'apport non-linéaire ne semblent pas très fiables à proximité de f_1 . Seules les valeurs au-delà de 100 Hz sont donc à interpréter. On peut alors conclure que dans la gamme de fréquence allant de 100 Hz à 4 kHz, la table d'harmonie de piano induit des non-linéarités dont l'amplitude est d'environ 4 – 5 % de celle du signal d'entrée. Cela permet de quantifier l'hypothèse selon laquelle la table d'harmonie du piano est linéaire.

4.3.4 Librairie Matlab *Hammerstein Toolbox*

Les outils développés à la section 4.2 et par Antonin Novák [Novák 2009, Novák et al. 2010a] permettent d'estimer rapidement un modèle non-linéaire correspondant à un système vibrant étudié. Une telle méthode est donc intéressante pour la communauté scientifique effectuant des mesures dynamiques car elle permet d'une part de tester les hypothèses de linéarité des systèmes étudiés et d'autre part d'avoir l'assurance que les contributions non-linéaires ont été correctement enlevées des réponses obtenues, même à fort niveau. D'autre part, une telle méthode peut aussi trouver de nombreuses applications dans la modélisation de systèmes faiblement non-linéaires (effets audio, haut-parleurs, ...).

Pour toutes ces raisons, une librairie Matlab intitulée *Hammerstein Toolbox* a été développée et rendue accessible en ligne sur le site *MatlabCentral*⁶. Cette librairie propose une implémentation des méthodes proposées à la section 4.2 et par [Novák 2009, Novák et al. 2010a]. Pour la méthode proposée à la section 4.2 différentes options sont offertes pour définir le filtre inverse (voir section 4.2.3).

4.4 Conclusion et perspectives

Ce chapitre traite de l'estimation et éventuellement de la correction des non-linéarités générées par les systèmes vibrants (haut-parleurs, pots vibrants, etc.) qui sont de façon générale au moins faiblement non-linéaires. La contribution majeure de ce chapitre est la mise au point et la validation d'une méthode originale, dérivée des travaux d'Angelo Farina [Farina 2000, Farina et al. 2001], permettant d'estimer rapidement un modèle de système non-linéaire correspondant à un système vibrant (voir Section 4.2). Puis, l'influence de différents paramètres sur la qualité des noyaux estimés est discutée (section 4.3.1), la méthode développée est comparée aux travaux d'Antonin Novák (voir section 4.3.2), le cas d'une chaîne de systèmes non-linéaires est étudié en regard des possibilités offertes par la méthode développée ici (voir section 4.3.3), et enfin la librairie Matlab disponible en ligne et proposant une implémentation des méthodes proposées en section 4.2 et par Novák et al. [2010a] est brièvement introduit (voir section 4.3.4).

Même si la mise au point et la validation de cette méthode répondent à l'interrogation "*Comment estimer et s'affranchir des non-linéarités générées par un système vibrant ?*" un certain nombre de questions de recherche restent encore ouvertes à la suite de ce travail :

- **Validité pratique des modèles de Hammerstein en cascade** : dans l'article présenté à la section 4.2, les systèmes faiblement non-linéaires étudiés en pratique (haut-parleur électro-dynamique et panneau

6. <http://www.mathworks.com/matlabcentral/>

vibrant) sont supposés être modélisables par des modèles de Hammerstein en cascade (*i.e.* avec des noyaux de Volterra diagonaux). Les résultats expérimentaux obtenus montrent que cette hypothèse conduit à des prédictions d'assez bonne qualité. Il semble donc que la majeure partie de l'énergie des noyaux de Volterra de ces systèmes réels soit contenue dans leur partie diagonale. Il serait cependant intéressant d'être capable de mesurer expérimentalement les noyaux de Volterra de ces systèmes jusqu'à l'ordre 3 ou 4 de façon à valider cette hypothèse de façon plus rigoureuse. Une approche complémentaire consisterait à utiliser un modèle physique non-linéaire de haut-parleur pour avoir des expressions analytiques des noyaux de Volterra [Klippel 1996a]. L'énergie contenue dans la partie diagonale de ces noyaux analytiques pourrait alors être comparée à l'énergie contenue hors de la diagonale.

- **Estimation de modèles plus riches** : la méthode présentée à la section 4.2 permet d'estimer des "*tranches de noyaux de Volterra*" relativement à un système non-linéaire donné (en pratique la *tranche* diagonale de chaque ordre de non-linéarité seulement). Récemment, Torras-Rosell & Jacobsen [2011] ont présenté un calcul analytique donnant la déconvolution de la sortie d'un système modélisé par une série de Volterra ayant reçu en entrée un sinus de fréquence variant exponentiellement. Ce calcul fait apparaître de façon cohérente avec l'article de la section 4.2 que les parties diagonales des noyaux de Volterra *polluent* la partie linéaire de la réponse et sont présentes aussi pour des temps $t < 0$. Ce calcul montre aussi que les parties non-diagonales *polluent* la partie linéaire et ont des composantes pour $t < 0$. Il est alors tentant de se demander s'il existe des signaux, multi-fréquentiels à un instant t donné, permettant l'estimation simultanée de plusieurs *tranches* d'un noyau de Volterra à un ordre donnée (et non plus seulement la *tranche* diagonale) en se basant sur une propriété de phase similaire à celle énoncée à la section 4.2. Ce type de signaux permettrait ainsi d'estimer des modèles plus riches que des cascades de modèles de Hammerstein.
- **Optimisation du signal d'entrée** : les résultats présentés à la section 4.2 et ceux de Novák et al. [2010a] font apparaître qu'il existe au moins deux signaux différents possédant la propriété de phase fondamentale nécessaire à l'estimation des modèles de Hammerstein en cascade. Il est alors intéressant de chercher mathématiquement la classe de signaux possédant cette propriété de phase fondamentale. Cela permettrait de choisir dans cette classe les signaux optimisant les résultats donnés par la méthode de la section 4.2 en fonction de certains critères (durée du signal d'entrée, rapport *signal sur bruit*, etc.).

Chapitre 5

Mesure dynamique des propriétés d'élasticité et de dissipation des constituants de plaques "sandwich"

Sommaire

5.1	Introduction	110
5.2	Étude qualitative du comportement dynamique d'une tranche de matériau sandwich symétrique	110
5.2.1	Définition d'un matériau de type "sandwich" symétrique	110
5.2.2	Types d'ondes existant dans un matériau sandwich symétrique	110
5.2.3	Modèles dynamiques élémentaires	112
5.2.4	Courbes de dispersion	114
5.2.5	Comportement asymptotique en basses fréquences	116
5.3	Détermination de paramètres d'élasticité et d'amortissement pertinents pour la modélisation de plaques "sandwich" épaisses	117
5.3.1	Introduction	117
5.3.2	Un modèle mécanique des panneaux "sandwich"	119
5.3.3	Estimation des paramètres modaux par une procédure de Rayleigh-Ritz étendue	124
5.3.4	L'analyse modale haute-résolution pour l'estimation des paramètres modaux	126
5.3.5	Procédure d'optimisation	129
5.3.6	Validation de la procédure d'estimation	131
5.3.7	Résultats expérimentaux	135
5.3.8	Conclusion	142
5.4	Discussion des hypothèses sur le rayonnement acoustique adoptées pour l'étude dynamique des panneaux	144
5.4.1	Hypothèse de <i>fluide léger</i>	144
5.4.2	Hypothèse de <i>faible amortissement</i>	148
5.4.3	Majoration des pertes par rayonnement	152
5.5	Conclusion	155

5.1 Introduction

Ce chapitre a pour objectif de présenter et de valider une méthode permettant l’estimation des paramètres mécaniques nécessaires à la description de la dynamique des panneaux de matériau sandwich composant le SMART-I². Cette méthode ne nécessite qu’une seule mesure dynamique, ponctuelle, réalisée sur une plaque taillée dans le matériau à décrire. Elle permet l’estimation des paramètres d’élasticité et d’amortissement du matériau testé, relatifs à une dynamique de plaque. Ces paramètres sont estimés par le biais d’une procédure mixte numérique/expérimentale consistant à mettre à jour les paramètres d’un modèle numérique en cherchant à minimiser l’écart entre les données obtenues expérimentalement et les données prédites par ce modèle.

Ce chapitre est divisé en trois parties. La première partie (section 5.2) propose une étude qualitative de la dynamique des matériaux sandwich et décrit les différents types d’ondes susceptibles de se propager dans un tel matériau. La seconde partie (section 5.3) est constituée d’un article paru dans le volume 330(25) du *Journal of Sound and Vibration*. Cet article présente la méthode proprement dite, une validation basée sur un modèle “*éléments finis*”, et les résultats expérimentaux obtenus sur différents échantillons. Enfin la section 5.4 discute le domaine de validité des différentes hypothèses sur le rayonnement acoustique qui ont été adoptées pour l’étude dynamique de la section 5.3.

5.2 Étude qualitative du comportement dynamique d’une tranche de matériau sandwich symétrique

5.2.1 Définition d’un matériau de type “sandwich” symétrique

Dans tout ce qui suit, un matériau sera dénommé “sandwich” s’il est composé de 3 couches distinctes, comme illustré sur la figure 5.1(a). Les deux couches extrêmes, ici beaucoup plus fines que la couche centrale sont appelées les “peaux”. La couche centrale est appelée le “cœur”. Un matériau sandwich est dit “symétrique” si au repos son plan médian est plan de symétrie du matériau [Nilsson 1990]. Dans ce cas de figure, les propriétés géométriques et mécaniques des peaux supérieure et inférieure sont les mêmes. Le matériau sandwich présenté à la figure 5.1(a) est symétrique.

5.2.2 Types d’ondes existant dans un matériau sandwich symétrique

Les notations spécifiées sur la figure 5.1(b) décrivent en partie le champ de déplacement d’une tranche de matériau sandwich symétrique. L’indice + se rapporte à la peau supérieure, et l’indice – à la peau inférieure. Au cours de cette première approche qualitative $w_{+,-}(x)$ désigne le déplacement transverse (*i.e.* selon z) d’un point de la peau supérieure ou inférieure situé à l’abscisse x lorsque la tranche de matériau est au repos. De la même façon, $u_{+,-}(x)$ désigne le déplacement longitudinal (*i.e.* selon x) d’un point de la peau supérieure ou inférieure initialement situé à l’abscisse x . Les déformations du cœur de la structure, qui n’est pas nécessairement homogène, ne sont pas décrites ici.

En décomposant les déplacements transverses qu’elles induisent sur chacune des peaux en la somme d’une composante symétrique et d’une composante antisymétrique par rapport au plan médian [Nilsson 1990, Sorokin 2002, Sorokin & Peake 2006], les deux catégories d’ondes suivantes peuvent être définies :

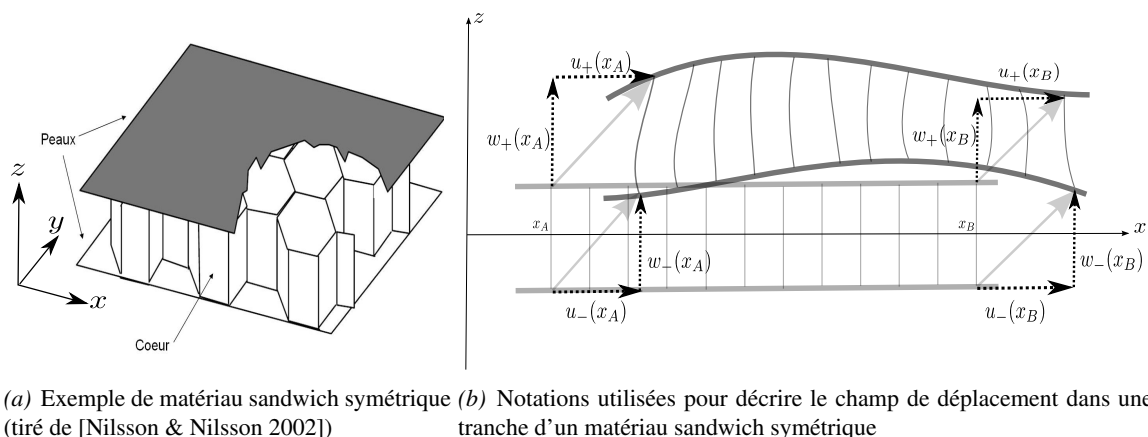


FIG. 5.1 – Matériau sandwich symétrique

1. Ondes “in-phase” : déplacements transverses en phase, $w_+(x) = w_-(x)$ et déplacements longitudinaux en opposition de phase, $u_+(x) = -u_-(x)$. La représentation de l'effet d'une onde de type *in-phase* sur une tranche de matériau sandwich symétrique initialement au repos est représenté sur la figure 5.2(a). Lors de leur propagation, ces ondes ne modifient pas l'épaisseur de la tranche de matériau sandwich et n'exercent pas d'efforts selon z dans le matériau sandwich. Les ondes de flexion ainsi que les ondes de cisaillement appartiennent à cette catégorie d'ondes.
2. Ondes “out-of-phase” : déplacements transverses en opposition de phase, $w_+(x) = -w_-(x)$ et déplacements longitudinaux en phase, $u_+(x) = u_-(x)$. La représentation de l'effet d'une onde *out-of-phase* sur une tranche de matériau sandwich symétrique initialement au repos est représenté sur la figure 5.2(b). Contrairement aux ondes *in-phase*, les ondes *out-of-phase* ne conservent pas l'épaisseur globale de la plaque.

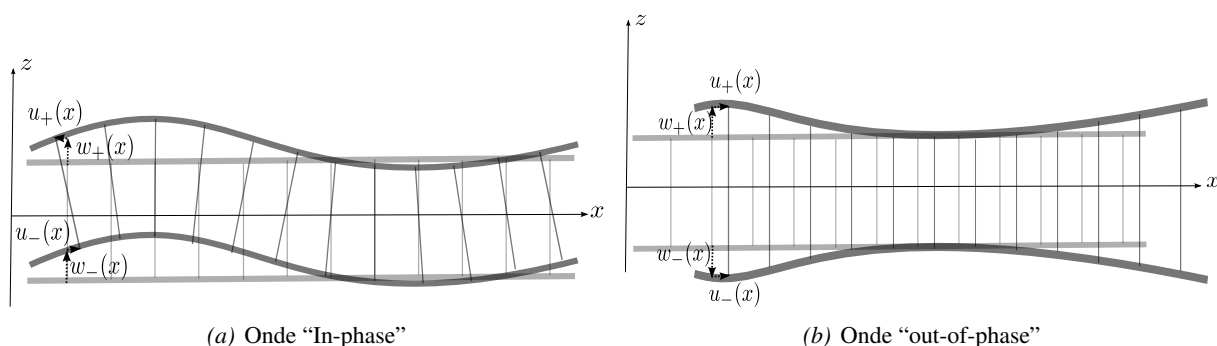


FIG. 5.2 – Illustration des deux catégories d'ondes pouvant se propager dans une tranche de matériau sandwich.

5.2.3 Modèles dynamiques élémentaires

Un nombre important de modèles analytiques plus ou moins détaillés et décrivant la dynamique des matériaux de type sandwich est disponible dans la littérature [Mead & Markus 1969, Mead 1982, Nilsson 1990, Nilsson & Nilsson 2002, Sorokin 2004, Yu & Cleghorn 2005, Hause & Librescu 2006, Liu & Zhao 2006, Kim 2007]. Pour une tranche de matériau sandwich symétrique et sous hypothèse de contraintes planes, il a été montré à l’aide de ces modèles que les ondes *in-phase* et les ondes *out-of-phase* définies à la section précédente étaient découplées les unes des autres [dos Santos et al. 1998, Caron 1999, Sorokin & Peake 2006]. Il est ainsi possible de modéliser en première approximation la propagation des ondes de chacune de ces catégories ondulatoires en utilisant des modèles élémentaires et indépendants les uns des autres plutôt qu’un unique modèle complexe [Sorokin 2004].

Ondes *in-phase*

Pour établir un modèle dynamique élémentaire de la propagation des ondes *in-phase*, l’hypothèse est faite par Nilsson & Nilsson [2002] que le déplacement transverse total des peaux $w(x, t) = w_+(x, t) = w_-(x, t)$ est dû aux effets combinés d’un mouvement de flexion du cœur et des peaux (déplacement angulaire β , voir figure 5.3(a)) et d’un mouvement de cisaillement du cœur (déplacement angulaire γ , voir figure 5.3(b)) de telle sorte que :

$$\frac{\partial w}{\partial x} = \gamma + \beta \quad (5.1)$$

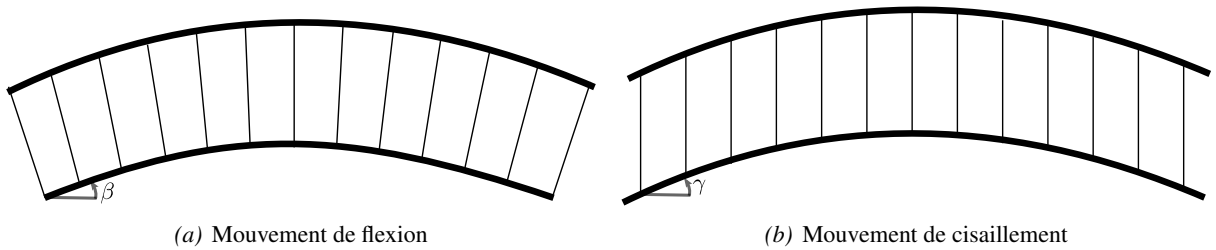


FIG. 5.3 – Déplacements induits par la flexion du cœur et des peaux (à gauche) et par le cisaillement du cœur (à droite) sur des sections droites d’une tranche de matériau sandwich symétrique.

Sous cette hypothèse, l’énergie potentielle totale de la tranche de matériau sandwich est alors la somme des contributions de la flexion des peaux et du cisaillement du cœur. De même l’énergie cinétique totale résulte des contributions des mouvements transverses et de rotation des sections droites de la tranche. L’utilisation du principe de Hamilton permet de déduire de ces énergies une équation de dispersion d’ordre 6 décrivant la propagation d’ondes planes *in-phase* dans la tranche [Nilsson & Nilsson 2002] :

$$2D_1D_hk^6 - 2ID_hk^4\omega^2 + (\mu D_1 + 2\mu D_h + IG_ch_c)k^2\omega^2 + G_ch_c(D_1k^4 + \mu\omega^2) + I\mu\omega^4 = 0 \quad (5.2)$$

avec

– E_c, E_p , modules de Young du cœur et de la peau,

- G_c , module de cisaillement du cœur,
- h_c, h_p , épaisseurs du cœur et de la peau,
- ρ_c, ρ_p , masses volumiques du cœur et de la peau,
- $D_l = E_c h_c^3/12 + E_p(h_c^2 + h_p/2 + h_c h_p^2 + 2h_p^3/3)$, rigidité totale de la tranche,
- $D_h = E_p h_p^3/12$, rigidité d'une peau seulement,
- $I = \rho_c h_c^3/12 + \rho_p(h_c^2 + h_p/2 + h_c h_p^2 + 2h_p^3/3)$, moment linéique d'inertie,
- $\mu = 2h_p \rho_p + h_c \rho_c$, masse surfacique de la tranche.

L'équation 5.2 possède trois couples de solutions. La fréquence caractéristique

$$f_t = \frac{1}{2\pi} \sqrt{\frac{G_c h_c}{I}} \quad (5.3)$$

délimite deux régimes ondulatoires distincts [Nilsson & Nilsson 2002] qui sont illustrés par la figure 5.4 :

- $f < f_t$: la déformation totale est la somme d'une onde propagative de flexion, d'un champ proche de flexion, et d'un champ proche de cisaillement (voir figure 5.4(a)).
- $f > f_t$: la déformation totale est la somme d'une onde propagative de flexion, d'une onde propagative de cisaillement, et d'un champ proche mixte de flexion et de cisaillement (voir figure 5.4(b)).

Ondes *out-of-phase* longitudinales

Le modèle décrivant la propagation d'une onde *out-of-phase* longitudinale est obtenu en posant $w(x, t) = 0$, $u(x, t) = u_+(x, t) = u_-(x, t)$ et en considérant uniquement les efforts élastiques longitudinaux [Sorokin 2004]. L'équation de dispersion qui découle de cette modélisation est :

$$D_{\text{long}} k^2 = \mu \omega^2 \quad \text{avec} \quad D_{\text{long}} = 2E_p h_p + E_c h_c \quad (5.4)$$

Cette équation de dispersion possède deux racines réelles qui traduisent la propagation d'ondes longitudinales dans deux directions opposées dans le matériau.

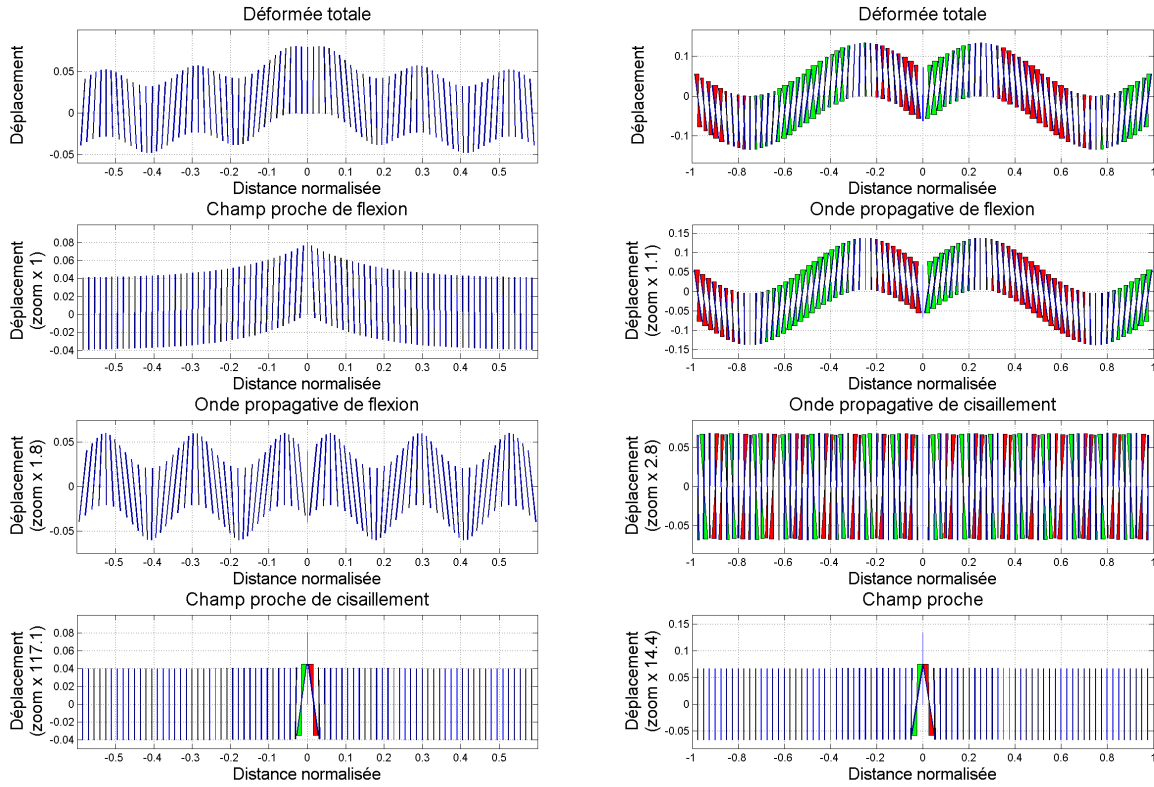
Ondes *out-of-phase* transverses

Le modèle décrivant la propagation des ondes *out-of-phase* transverses est obtenu en ignorant les déplacements longitudinaux $u_+(x, t) = u_-(x, t) = 0$, l'inertie du cœur et en supposant $w(x, t) = w_+(x, t) = -w_-(x, t)$ [Sorokin 2004]. De ce fait, le comportement du cœur est réduit à celui d'une raideur surfacique équivalente, distribuée, de valeur $K_{\text{eq}} = E_c/h_c$. L'équation de dispersion décrivant la propagation d'onde planes *out-of-phase* transverses dans la tranche de matériau sandwich est alors donnée par :

$$D_h k^4 - \rho_p h_p \omega^2 + 2K_{\text{eq}} = 0 \quad (5.5)$$

Cette équation est d'ordre 4 et décrit la propagation dans des sens opposés de deux types d'onde . En dessous de la fréquence

$$f_d = \frac{1}{2\pi} \sqrt{\frac{K_{\text{eq}}}{\rho_p h_p}} \quad (5.6)$$


 (a) Fréquence normalisée $f = 0.4f_t$

 (b) Fréquence normalisée $f = 2f_t$

 FIG. 5.4 – Déformation totale d’une tranche monodimensionnelle infinie de matériau sandwich symétrique excitée transversalement à la distance normalisée 0 et à la fréquence f .

La déformation totale est décomposée en la somme de ses différentes composantes. Les distances et déplacements sont normalisés par la plus grande des trois longueurs d’onde en présence. Les barres bleues symbolisent les sections droites de la tranche et sont verticales au repos. La part de la rotation des sections droites de la tranche due au cisaillement du cœur est représentée par les triangles verts (sens horaire) et rouges (sens anti-horaire). Cette simulation a été réalisée d’après la modélisation proposée par Backstrom & Nilsson [2007] et avec $h_c = 5$ cm, $h_p = 5$ mm, $\rho_p = 1760$ kg/m³, $\rho_c = 130$ kg/m³, $E_p = 16.7$ GPa, $E_c = 0.13$ GPa, $G_c = 0.05$ GPa, et donc $f_t = 2100$ Hz.

les nombres d’onde solution en (w_+) de l’Eq. (5.5) comportent toujours une partie réelle et une partie imaginaire non nulles. Les deux ondes associées sont donc toujours évanescentes. Pour $f > f_d$, l’une d’entre elles devient purement propagative tandis que l’autre devient purement évanescence.

5.2.4 Courbes de dispersion

La figure 5.5 donne les courbes de dispersion correspondant à chacun des couples d’ondes décrits à la section 5.2.3. La courbe *in-phase 1* correspond au champ proche de flexion pour $f < f_t$ et à une onde propagative où la déformation est dominée par la flexion pour $f > f_t$. La courbe *in-phase 2* correspond à une onde propagative où la déformation est dominée par la flexion pour $f < f_t$ qui se transforme en une onde propagative où la déformation est dominée par le cisaillement pour $f > f_t$. La courbe *in-phase 3* représente le champ proche mixte de cisaillement et de flexion. La courbe *out-of-phase longitudinale re-*

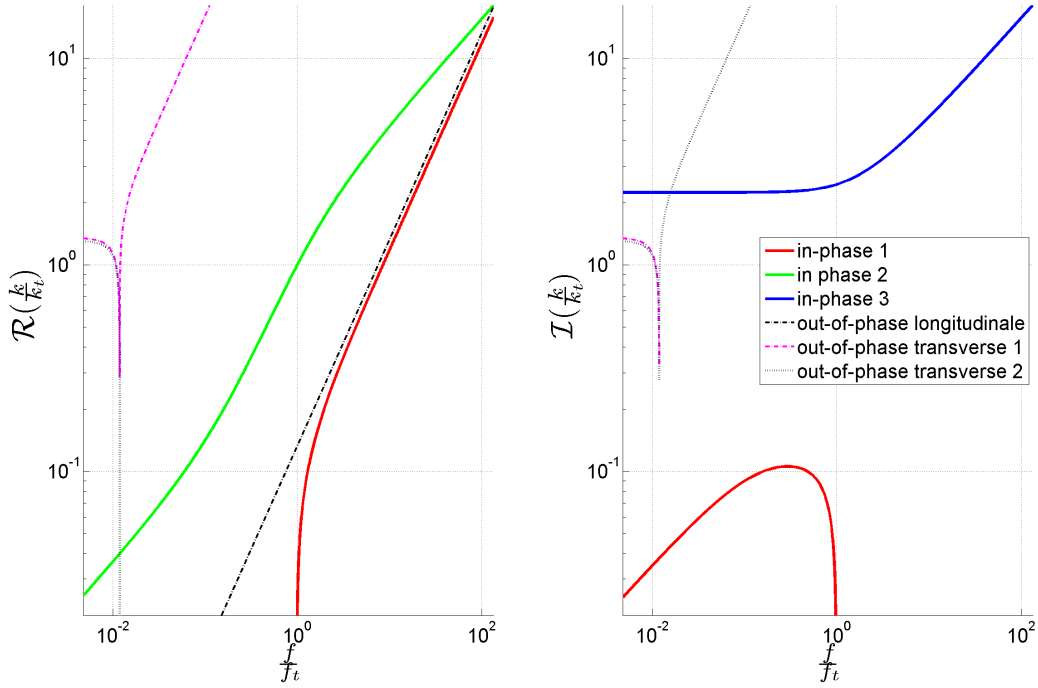


FIG. 5.5 – Courbes de dispersion obtenues par la superposition des modèles élémentaires établis à la section 5.2.3 pour une tranche de matériau sandwich symétrique.

Les fréquences sont normalisées par rapport à $f_t = 2100$ Hz. Les nombres d'onde sont normalisés par le nombre d'onde de l'onde de flexion propagative à la fréquence de normalisation f_t . Les paramètres mécaniques et géométriques utilisés pour tracer cette courbe sont identiques à ceux utilisés pour générer la figure 5.4.

présente les ondes propagatives longitudinales existant dans le matériau. La courbe *out-of-phase transverse 1* décrit une onde avec une composante évanescence pour $f < f_d$ et une onde propagative pour $f > f_d$. La courbe *out-of-phase transverse 2* décrit une onde avec une composante évanescence pour $f < f_d$ et une onde purement évanescence pour $f > f_d$.

L'étude de la figure 5.5 révèle les différents ordres de grandeurs en jeu dans l'étude de la dynamique d'une tranche de matériau sandwich symétrique. En premier lieu, il existe plusieurs ordres de grandeur de différence entre les nombres d'onde associés aux ondes *in-phase 1* & 2 et *out-of-phase longitudinale* et ceux associés aux ondes *in-phase 3* et *out-of-phase transverses 1* & 2. Cela signifie que les longueurs d'ondes associées à ces déformations sont bien plus petites que celles associées aux ondes *in-phase 1* & 2 et *out-of-phase longitudinale*. D'autre part, on note aussi que la fréquence f_t décrivant la transition entre deux modes comportementaux pour les ondes *in-phase 1* & 2 est bien plus grande que la fréquence f_p décrivant l'apparition d'onde propagatives transverses.

En réalité, les modèles plus complexes du champ de déplacement démontrent l'existence d'un nombre croissant de courbes de dispersion au fur et à mesure que la fréquence augmente, et non pas seulement 6 comme le suggère la figure 5.5. D'après Sorokin [2004], les différents modèles dynamiques élémentaires établis précédemment sont cependant de bonnes approximations des 6 premières courbes de dispersion.

5.2.5 Comportement asymptotique en basses fréquences

Dans cette section, seules les ondes *in-phase 1* & 2 seront étudiées pour $f < f_t$. Sur la figure 5.4 sont représentées les courbes de dispersion correspondant aux ondes *in-phase 1* & 2 pour une tranche de matériau sandwich symétrique et les courbes de dispersion correspondant à des modèles de tranches simplifiés.

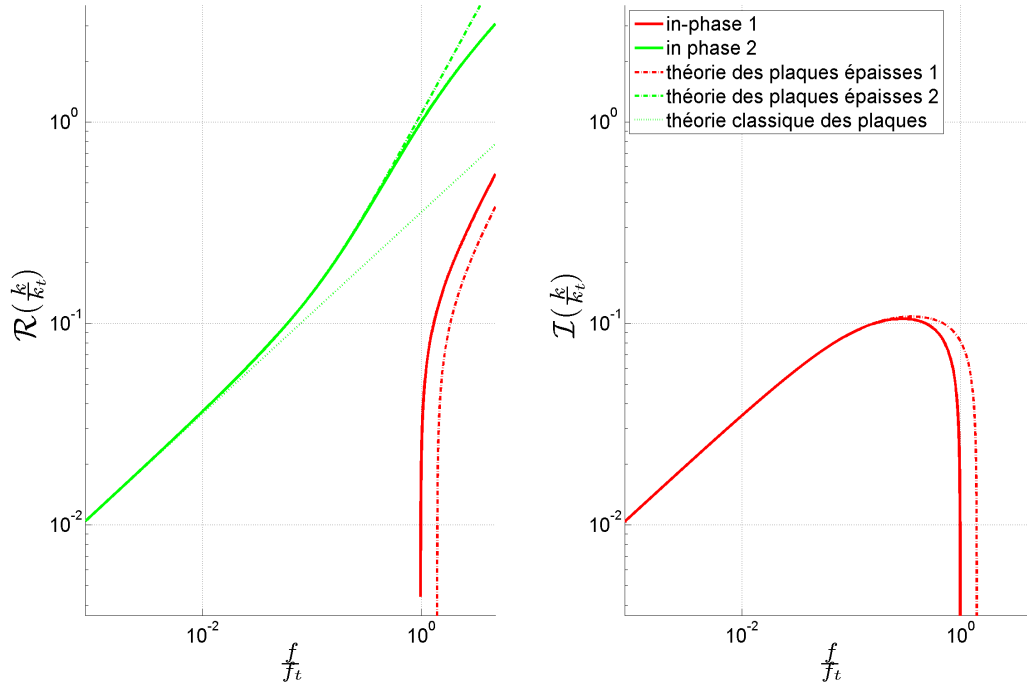


FIG. 5.6 – Courbes de dispersion obtenues par la superposition de différents modèles de plaques.

Les fréquences sont normalisées par rapport à $f_t = 2100$ Hz. Les nombres d’onde sont normalisés par le nombre d’onde de l’onde de flexion propagative à la fréquence de normalisation f_t . Les paramètres mécaniques et géométriques utilisés pour tracer cette courbe sont identiques à ceux utilisés pour générer la figure 5.4.

Parmi ces modèles simplifiés figure la théorie classique des plaques de Cremer et al. [2005], qui considère une tranche faite d’un matériau homogène et qui néglige les phénomènes de cisaillement dans l’épaisseur de la tranche. L’étude de la figure 5.6 révèle que la courbe de dispersion correspondant aux ondes *in-phase 2* qui est prédite par le modèle élémentaire présenté à la section 5.2.3 converge asymptotiquement en basse fréquence vers celle donnée par la théorie classique des plaques. L’approximation consistant à supposer que la tranche de sandwich se comporte comme une plaque épaisse est donc valide pour $f < 0.1f_t$. Dans ce cas, la rigidité de la tranche vue par les ondes de flexion s’y propageant est D_1 et correspond à la rigidité totale de la tranche. Cependant, la théorie classique des plaques ne permet pas de modéliser la courbe de dispersion correspondant aux ondes *in-phase 1*.

Un modèle plus raffiné consiste alors à considérer les hypothèses de Reissner-Mindlin [Cremer et al. 2005] (théorie des plaques épaisses) : les effets linéaires du cisaillement dans l’épaisseur de la tranche ainsi que l’inertie due à la rotation des sections droites sont à présent pris en compte. La tranche est toujours considérée comme homogène. L’étude de la figure 5.6 révèle que les courbes de dispersion correspondant

aux ondes *in-phase 1 et 2* qui sont prédites par le modèle élémentaire présenté à la sous-section 5.2.3 sont proches de celles données par le modèle de plaque épaisse quasiment jusqu'à $f \simeq f_t$. Dans ce cas, la rigidité en flexion de la tranche vue par les ondes s'y propageant est D_1 , la rigidité totale de la tranche et le module de cisaillement effectif de la tranche est G_c , le module de cisaillement du cœur. La théorie des plaques épaisses permet donc d'approximer correctement les ondes *in-phase 1 et 2* pour $f < f_t$.

5.3 Détermination de paramètres d'élasticité et d'amortissement pertinents pour la modélisation de plaques "sandwich" épaisses

Marc Rébillat, Xavier Boutillon

Cette section reproduit l'article intitulé "*Measurement of relevant elastic and damping material properties in sandwich thick-plates*" paru dans le volume 330(25) du *Journal of Sound and Vibration* [Rébillat & Boutillon 2011a]. Les différentes parties composant cet article ont fait l'objet de présentations lors de l'ICEM à Poitiers [Rébillat & Boutillon 2010b], lors de l'ISMA à Louvain [Rébillat & Boutillon 2010a] et lors du CSMA à Giens [Rébillat & Boutillon 2011b]. Les hypothèses sur le rayonnement acoustique adoptées pour l'étude dynamique des panneaux décrite dans cet article seront discutées à la section 5.4.

Abstract

An easy-to-implement method to measure relevant elastic and damping properties of the constituents of a sandwich structure, possibly with a heterogeneous core, is proposed. The method makes use of a one-point dynamical measurement on a thick plate. The hysteretic model for each (possibly orthotropic) constituent is written generically as " $E(1 + j\eta)$ " for all mechanical parameters. The estimation method of the parameters relies on a mixed experimental / numerical procedure. The frequencies and dampings of the natural modes of the plate are obtained from experimental impulse responses by means of a high-resolution modal analysis technique. This allows for considerably more experimental data to be used. Numerical modes (frequencies, dampings, and modal shapes) are computed by means of an extended Rayleigh-Ritz procedure under the "light damping" hypothesis, for given values of the mechanical parameters. Minimising the differences between the modal characteristics yields an estimation of the values of the mechanical parameters describing the hysteretic behaviour. A sensitivity analysis assess the reliability of the method for each parameter. Validations of the method are proposed by (a) applying it to virtual plates on which a finite-element model replaces the experimental modal analysis, (b) some comparisons with results obtained by static mechanical measurements, and (c) by comparing the results on different plates made of the same sandwich material.

5.3.1 Introduction

For plates having a mechanical function, sandwich structures, with possibly a heterogeneous core, are often preferred to a homogeneous constitutive material because they can be made lighter. However, the relevant mechanical properties of the sandwich as a whole or even of its individual components may be

difficult to predict accurately, particularly when heterogeneous cores are used or if damping is considered. Here “*relevant*” refers to the parameters that matter in the plate dynamics (see Sec. 5.3.2).

In this paper, a method for estimating the complex moduli of elasticity of the constituents of sandwich structures having a heterogeneous core based on one dynamical test on a plate is proposed. The proposed mixed experimental / numerical procedure (for an introduction to such procedures, see for example [Ayorinde & Yu 2005, Lauwagie et al. 2006]) is based on the thick-plate model and frequency-independent mechanical properties of each constituent of the sandwich. It is intended to be easier to implement or to yield more parameters than other methods.

Using structural vibrations for the estimation of homogeneous material parameters is a widespread technique. Compared to static measurement campaigns which must be performed on a number of isolated samples of the sandwich components, dynamic tests leave the sandwich structure untouched and can be performed on one single panel of more or less arbitrary dimensions.

The problem of the estimation of solely *elasticity* parameters of a homogeneous material using plate vibrations has been widely addressed in a *thin-plate* context [Deobald & Gibson 1988, Moussu & Nivoit 1993, Grediac et al. 1998, Hwang & Chang 2000, Muthurajan et al. 2004, Ayorinde & Yu 2005, Lauwagie et al. 2006] and in a *thick-plate* context [Soares et al. 1993, Araujo et al. 1996, Ayorinde 1996, Lai & Ip 1996, Frederiksen 1997b,c,a, 1998, Bledzki et al. 1999, Cunha & Piranda 1999, Araujo et al. 2000, Gagneja et al. 2001, Cugnoni et al. 2004, 2007]. In a *thin-plate* context, only *in-plane* parameters can be estimated. Using a *thick-plate* yields some of the so-called *out-of-plane* parameters.

The problem of estimating *elasticity and damping* parameters of homogeneous materials by using point measurements [Mac Intire & Woodhouse 1988, de Visscher et al. 1997, Talbot & Woodhouse 1997, Qian et al. 1997] has retained some attention in a *thin-plate* context. In a *thick-plate* context, methods involving *full-field* measurements are currently available [Pagnacco et al. 2007, Matter et al. 2009, Giraudeau et al. 2010], but they are very time-consuming or need sophisticated equipment. Recently, some efforts have however been done to overcome these limitations by using point measurements instead of full-field measurements [Araujo et al. 2009b]. However, the experimental data used by these authors severely limits the possibilities of estimation, as discussed below.

For heterogeneous materials, such as sandwich materials, the literature addresses mainly the estimation of *elasticity* parameters, using either beams [Saito et al. 1997, Shi et al. 2006, Barbieri et al. 2010] or plates [Bastos et al. 2002, Ichchou et al. 2008]. A few studies have been devoted in the last years to the problem of estimating their damping properties in beams [Shi et al. 2006, Barbieri et al. 2010], in plates by means of sophisticated full-field measurements [Berthaut et al. 2003, Matter et al. 2011], and quite recently in plates by single point measurements [Barkanov et al. 2009, Araujo et al. 2010]. Since none of these studies present a sensitivity analysis, the validity of the model they used for damping is difficult to assess. The following reasoning explains how we access to significantly more experimental data than previous studies. The sensitivity analysis (Sec. 5.3.7) shows that these are determinant for some estimated parameters.

Exploiting the vibrations of sandwich panels with heterogeneous cores requires that the panel dimensions meet several conditions. In order to consider the sandwich core as homogeneous in the in-plane directions up to a given frequency f , the corresponding wavelength λ must contain at least 50 cells [Burton & Noor 1997]. For a typical cell side-length s_{cell} and height h , this implies that the panel’s dimensions are such

that $l_{x,y} > \lambda > 50 s_{\text{cell}}$. Moreover *out-of-plane* elasticity and damping parameters can be estimated only on thick-enough plates. In brief, the panel must be large enough and the observed dynamics must include high-enough modes, within the limit of a plate model. Due to the intrinsic dissipation of materials, high-order modes may be impossible to characterise with methods based on the Fourier-Transform (FT): those are limited to modal overlaps of $\mu \simeq 30\%$ in most implementations [Berthaut et al. 2003] (see Sec. 5.3.4 for the definition of μ). The high-resolution modal analysis (HRMA) technique recently developed by Ege *et al.* [Ege et al. 2009] is a successful substitute to the FT-based modal analysis techniques up to significantly higher modal overlap values.

From a methodological point of view, the present work is an extension to heterogeneous and thick plates of the work done by De Visscher *et al.* [de Visscher et al. 1997] on homogeneous thin plates. The proposed method is performed on large sandwich panels which meet the above conditions. Up to $\simeq 40$ modes are extracted by means of the HRMA. This is considerably more than the 6 or 12 modes used in [Araujo et al. 2009b] for the determination of 6 elastic parameters. It will be shown in Sec. 5.3.7 that high-frequency modes are indeed necessary for a reliable estimation of some of these parameters. As far as damping parameters are considered, experimental data used in [Araujo et al. 2009b] do not allow their reliable determination (as stated by the authors and confirmed by the sensitivity analysis in Sec. 5.3.7).

The mixed experimental / numerical procedure yielding the *in-plane* and most *out-of-plane* elasticity and damping parameters (complex elasticity moduli) of the constituents of the sandwich is schematically presented in Fig. 5.7. The analytical model of the sandwich panel is presented in Sec. 5.3.2. Based on this model, the numerical modal frequencies f_n^{Num} and dampings α_n^{Num} are derived by means of an extended Rayleigh-Ritz procedure (Sec. 5.3.3). The experimental protocol and the extraction method that yield the experimental modal frequencies f_n^{XP} and dampings α_n^{XP} are presented in Sec. 5.3.4. Given the numerical and experimental data, the optimisation procedure that estimates elasticity and damping parameters of the constituents of the sandwich material is detailed in Sec. 5.3.5. The procedure is validated in Sec. 5.3.6 by means of a finite-element analysis. Measurements performed on 3 real plates are shown in Sec. 5.3.7 which provides additional validation insight as well.

5.3.2 A mechanical model of sandwich panels

In order to access the modal dampings and frequencies of a sandwich panel, an adapted mechanical model is needed. In this section, such a complex structure is modelled as an equivalent thick-plate under the Reissner-Mindlin hypothesis. A frequency-independent model for the materials composing the sandwich is also described. Notations used in this section are summarized by Tabs. B.1, B.2 and B.3 in the Appendix B.4.

Hypothesis

The sandwich panel consists in two identical skins and a core (Fig. 5.8). The thicknesses of the core, skins, and panel are h^c , h^s , and $h = h^c + 2h^s$ respectively. In the following, "panel" designates the physical structure whereas "plate" refers to the idealised structure made out of the equivalent homogeneous material. The following hypotheses are made on the panel and plate:

- Displacements are small so that the materials and structures behave linearly.

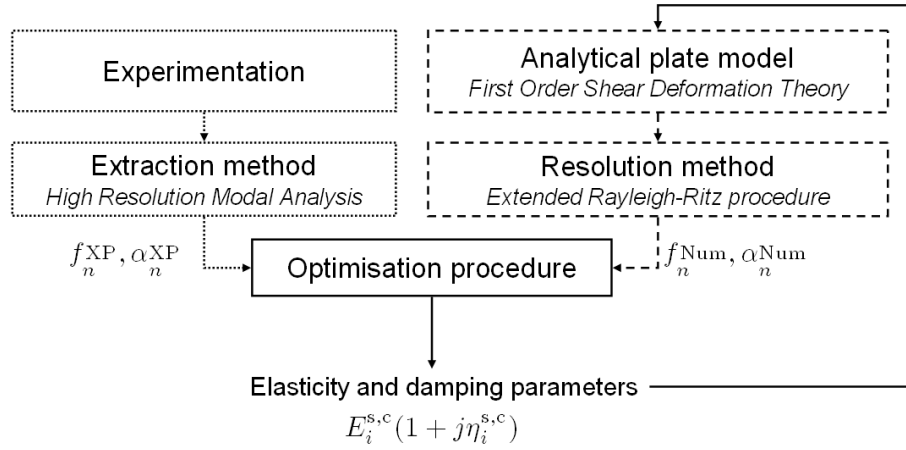


Figure 5.7: Overview of the proposed mixed experimental / numerical procedure: modal characteristics derived experimentally and numerically are compared; their differences tend to zero when the correct values of the mechanical parameters are reached.

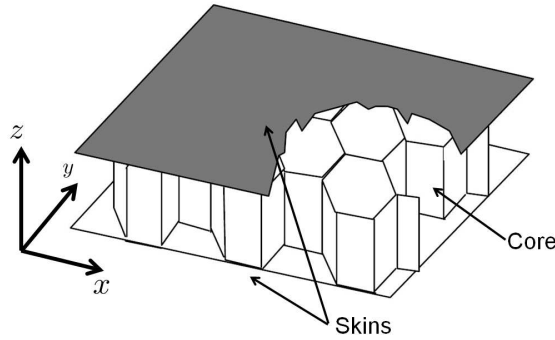


Figure 5.8: Geometry of the sandwich plate.

- Only flexural waves of frequencies far from the delamination frequency (the frequency of the first transverse mode of the panel) are considered.
- The plate is considered to follow the Reissner-Mindlin approximations (thick-plate: first order shear deformation theory, FSDT), with no direct strain in the z -direction. As pointed out by references [Yu & Cleghorn 2005, Liu & Zhao 2006, Kim 2007], the accuracy of the thick-plate model to describe the dynamics of sandwich plates having a soft core depends mainly on the thickness to length ratio and on the skins Young modulus to core shear modulus ratio: $\frac{E_{x,y}^s}{G_{yz,xz}^c}$. Generally, the thicker and the softer the core, the less the thick-plate model is appropriate for modelling purposes as compared to higher-order models. In [Liu & Zhao 2006, Fig. 4], a comparison between thick-plate theory (FSDT) and high-order shear deformation theory (HSDT) is performed on the first modal frequency of a sandwich panel with a soft core. At least for orders of magnitude considered, it could be extended to a higher modal frequency of wavelength λ by considering the ratio $\frac{2h}{\lambda}$ instead of $\frac{h}{l_{x,y}}$ (the “ h ” written here is defined in Fig. 5.8 and corresponds to the writing “ $2h$ ” in [Liu & Zhao 2006]). According to [Liu

& Zhao 2006, Fig. 4(c) and 4(d)], the difference between FSTD and HSDT does not exceed 5% for $\frac{E_{x,y}^s}{G_{yz,xz}^c} \simeq 100$ and a $\frac{2h}{\lambda}$ ratio less than 0.08. The plates that are considered in this paper exhibit a $\frac{E_{x,y}^s}{G_{yz,xz}^c} \simeq 100$ ratio and the highest modes under consideration are characterized by $\frac{2h}{\lambda} \simeq 0.08$. A thick-plate model based on FSTD is thus appropriate enough in the present study.

- The wavelengths λ include at least 50 cells. According to Burton *et al.* [Burton & Noor 1997], this ensures that errors on the modal frequencies of the plate (with a homogeneous equivalent core) are less than 2% when compared to those of the panel as computed by various finite-elements models.

The skin and core materials are considered as homogeneous, orthotropic in the x – and y – directions.

The formalism chosen for describing the hysteretic behaviour is that of complex moduli $\underline{E} = E(1 + j\eta)$ which does not depend on the frequency (see the model of materials in section 5.3.2). The Young's and shear moduli and the Poisson coefficient of the core are \underline{E}_x^c , \underline{E}_y^c , \underline{E}_z^c , \underline{G}_{xy}^c , \underline{G}_{xz}^c , \underline{G}_{yz}^c , $\underline{\nu}_{yx}^c$, $\underline{\nu}_{xz}^c$, $\underline{\nu}_{yz}^c$ and $\underline{\nu}_{xy}^c$. The same parameters for the skins are denoted by the index "s". The properties of the homogeneous material equivalent to the whole sandwich are denoted by the index "H". Additional symmetry relationships are given in section 5.3.2.

The following hypotheses are made on the sandwich panel:

- The sandwich panel is symmetric with respect to its mid-plane.
- Skins are thin compared to the core and the core is softer than the skins so that shear stress in the skins can be ignored: $\frac{h^s}{G_{xz}^s} \ll \frac{h^c}{G_{xz}^c}$ (and the same in the y direction).
- The core is very soft ($E_x^c \ll E_x^s$, $E_y^c \ll E_y^s$ and $G_{xy}^c \ll G_{xy}^s$). Given the generic expression of the moduli of the homogeneous equivalent material $E^H = \left(\frac{h^c}{h}\right)^3 E^c + \left[1 - \left(\frac{h^c}{h}\right)^3\right] E^s$, this ensures that all in-plane stress in the plate are entirely due to those in the skins.

According to these hypotheses, there is no stress associated with $\underline{E}_z^{c,s,H}$, $\underline{\nu}_{xz}^{c,s,H}$, $\underline{\nu}_{yz}^{c,s,H}$, \underline{G}_{xz}^s , \underline{G}_{yz}^s , \underline{E}_x^c , \underline{E}_y^c , \underline{G}_{xy}^c , $\underline{\nu}_{xy}^c$, $\underline{\nu}_{yx}^c$ which are ignored in what follows. The relevant remaining mechanical parameters describing such a plate are thus [Matter et al. 2011]:

- *In-plane* parameters (bending of the skins) : \underline{E}_x^s , \underline{E}_y^s , \underline{G}_{xy}^s , $\underline{\nu}_{yx}^s$, $\underline{\nu}_{xy}^s$.
- *Out-of-plane* parameters (shearing in the core) : \underline{G}_{xz}^c , \underline{G}_{yz}^c (excluding \underline{E}_z^c).

In the rest of the article, it will thus be understood that "*elastic parameters*" means "*relevant elastic parameters for the dynamics of thick-plates*".

These hypotheses are generally fulfilled in common sandwich panels. Typical orders of magnitude for parameters are:

$$\begin{cases} h^s/h^c \simeq 10^{-1} \\ E_x^c/E_x^s \simeq E_y^c/E_y^s \simeq G_{xy}^c/G_{xy}^s \simeq 10^{-4} \end{cases} \quad (5.7)$$

Model of the materials

The damping of plate vibrations has different origins. In the present study, it is assumed that panels vibrate below their coincidence acoustical frequencies [Wallace 1972]. Consequently, damping due to acoustical radiation in surrounding air is very small compared to the structural damping [Chaigne & Lam-bourg 2001]. Among the different structural damping models, the standard hysteretic model (which is

frequency-independent, see for example [Pritz 1998, Muravskii 2004]) has been retained. The relationship between the stress ϵ^γ and the strain σ^γ in each γ -material ($\gamma = s, c$, or H) involves 7 complex numbers and can be written, to first order in η^γ as:

$$\sigma^\gamma = \begin{bmatrix} E_x^\gamma(1 + j\eta_x^\gamma) & \nu_{yx}^\gamma E_x^\gamma[1 + j(\eta_{yx}^\gamma + \eta_x^\gamma)] & 0 & 0 & 0 \\ \nu_{xy}^\gamma E_y^\gamma[1 + j(\eta_{xy}^\gamma + \eta_y^\gamma)] & E_y^\gamma(1 + j\eta_y^\gamma) & 0 & 0 & 0 \\ 0 & 0 & G_{xz}^\gamma(1 + j\eta_{xz}^\gamma) & 0 & 0 \\ 0 & 0 & 0 & G_{yz}^\gamma(1 + j\eta_{yz}^\gamma) & 0 \\ 0 & 0 & 0 & 0 & G_{xy}^\gamma(1 + j\eta_{xy}^\gamma) \end{bmatrix} \epsilon^\gamma \quad (5.8)$$

The symmetry of the strain/stress relation adds the following relationships $\nu_{xy}^\gamma E_y^\gamma = \nu_{yx}^\gamma E_x^\gamma$ and $\eta_\nu^\gamma = \eta_{\nu_{xy}}^\gamma + \eta_y^\gamma = \eta_{\nu_{yx}}^\gamma + \eta_x^\gamma$ which leaves 12 independent real parameters to be identified for each material (24 altogether).

Equivalent thick-plate

Under the hypothesis and for the orders of magnitude given in section 5.3.2, the sandwich panel behaves in the low frequency range like a homogeneous thick-plate [Nilsson & Nilsson 2002]. The thickness of the plate is chosen to be h . Its mechanical properties are given in Eq. (5.9) and (5.10) as functions of the mechanical and geometrical properties of the skins and the core ¹.

$$\begin{cases} E_x^H = E_x^s \left[1 - \left(\frac{h^c}{h} \right)^3 \right] & E_y^H = E_y^s \left[1 - \left(\frac{h^c}{h} \right)^3 \right] & \nu_{xy}^H = \nu_{xy}^s \\ G_{xy}^H = G_{xy}^s \left[1 - \left(\frac{h^c}{h} \right)^3 \right] & G_{xz}^H = G_{xz}^c & G_{yz}^H = G_{yz}^c \end{cases} \quad (5.9)$$

$$\begin{cases} \eta_x^H = \eta_x^c \frac{E_x^c}{E_x^s} \left(\frac{h^c}{h} \right)^3 + \eta_x^s \left[1 - \left(\frac{h^c}{h} \right)^3 \right] & \eta_y^H = \eta_y^c \frac{E_y^c}{E_y^s} \left(\frac{h^c}{h} \right)^3 + \eta_y^s \left[1 - \left(\frac{h^c}{h} \right)^3 \right] \\ \eta_{xy}^H = \eta_{xy}^c \frac{G_{xy}^c}{G_{xy}^s} \left(\frac{h^c}{h} \right)^3 + \eta_{xy}^s \left[1 - \left(\frac{h^c}{h} \right)^3 \right] & \eta_{xz}^H = \eta_{xz}^c & \eta_{yz}^H = \eta_{yz}^c & \eta_\nu^H = \eta_\nu^s \end{cases} \quad (5.10)$$

The 12 independent real parameters $\{E_x^H, \eta_x^H, E_y^H, \eta_y^H, G_{xy}^H, \eta_{xy}^H, G_{xz}^H, \eta_{xz}^H, G_{yz}^H, \eta_{yz}^H, \nu_{xy}^H, \eta_\nu^H\}$ are to be estimated. Their knowledge yields the elastic and damping properties of each layer of the sandwich panel provided that the 12-equation system formed by Eqs. (5.9) and (5.10) is invertible. A sufficient condition is:

$$\eta_x^c \frac{E_x^c}{E_x^s} \ll \eta_x^s \quad \eta_y^c \frac{E_y^c}{E_y^s} \ll \eta_y^s \quad \eta_{xy}^c \frac{G_{xy}^c}{G_{xy}^s} \ll \eta_{xy}^s \quad (5.11)$$

1. Étant donné qu’une structure tri-couche complexe est modélisée ici avec un simple modèle de plaque épaisse homogène, il est alors probable que les modules d’élasticité du modèle simplifié dépendent de la fréquence. Comme le montrent Backstrom & Nilsson [2007], la dépendance fréquentielle des modules d’élasticité peut être négligée en dessous d’une certaine fréquence f_t dépendant de la géométrie et des propriétés mécaniques du matériau étudié. Dans le cas du matériau étudié ici, $f_t \simeq 14$ kHz. Compte tenu du domaine de fréquence abordé, il est donc justifié de négliger les variations en fréquences des modules d’élasticité du modèle simplifié.

since $\frac{E_x^c}{E_x^s} \ll 1$, $\frac{E_y^c}{E_y^s} \ll 1$, and $\frac{G_{xy}^c}{G_{xy}^s} \ll 1$ (see section 5.3.2). This condition is not satisfied only if the η^c -coefficients are several orders of magnitude larger than the η^s -ones. This is not the case here and rarely the case in general². Consequently, the estimation of E_x^H , etc ... yields an estimation of the mechanical properties of the skin and core materials.

Potential, kinetic and dissipated energies in the equivalent thick-plate

Within the frame of the first order Reissner-Mindlin theory [Cremer et al. 2005, Chap. 3], the displacements $\{u, v, w\}$ in the $\{x, y, z\}$ -directions respectively can be written within a good approximation (see below) as:

$$u(x, y, z) = -z\Phi_x(x, y) \quad v(x, y, z) = -z\Phi_y(x, y) \quad w(x, y, z) = w_0(x, y) \quad (5.12)$$

The potential energy of the plate is:

$$\begin{aligned} U &= \frac{1}{2} \iiint_V (\boldsymbol{\sigma}^H)^T \boldsymbol{\epsilon}^H d\tau \\ &= \frac{1}{2} \iint_S \left[D_1 \left(\frac{\partial \Phi_x}{\partial x} \right)^2 + D_2 \left(\frac{\partial \Phi_x}{\partial x} \frac{\partial \Phi_y}{\partial y} \right) + D_3 \left(\frac{\partial \Phi_y}{\partial y} \right)^2 + \dots \right. \\ &\quad D_4 \left(\Phi_y^2 - 2\Phi_y \frac{\partial w_0}{\partial y} + \left(\frac{\partial w_0}{\partial y} \right)^2 \right) + D_5 \left(\Phi_x^2 - 2\Phi_x \frac{\partial w_0}{\partial x} + \left(\frac{\partial w_0}{\partial x} \right)^2 \right) + \dots \\ &\quad \left. D_6 \left(\left(\frac{\partial \Phi_x}{\partial y} \right)^2 + 2 \frac{\partial \Phi_x}{\partial y} \frac{\partial \Phi_y}{\partial x} + \left(\frac{\partial \Phi_y}{\partial x} \right)^2 \right) \right] dx dy \end{aligned} \quad (5.13)$$

with

$$\begin{aligned} D_1 &= \frac{E_x^H h^3}{12(1 - \nu_{xy}\nu_{yx})} & D_2 &= \frac{\nu_{xy} E_y^H h^3}{6(1 - \nu_{xy}\nu_{yx})} & D_3 &= \frac{E_y^H h^3}{12(1 - \nu_{xy}\nu_{yx})} \\ D_4 &= 2\kappa_{yz}^2 h G_{yz}^H & D_5 &= 2\kappa_{xz}^2 h G_{xz}^H & D_6 &= \frac{G_{xy}^H h^3}{6} \end{aligned} \quad (5.14)$$

The shear correction factors κ_{yz}^2 and κ_{xz}^2 account for the fact that Eq. (5.12) is an approximation: the (functional) angles Φ_x and Φ_y depend lightly on z and sections of the plate do not remain plane in the flexural deformation. The values $\kappa_{yz} = \kappa_{xz} = 1$ have been chosen according to the recommendations of Birman [2002] for sandwich panels.

By definition, the fraction of energy lost during one cycle \mathcal{T} is:

$$\Delta U = - \int_{\mathcal{T}} \left[\iiint_V (\boldsymbol{\sigma}^H)^T \frac{\partial \boldsymbol{\epsilon}^H}{\partial t} d\tau \right] dt \quad (5.15)$$

Based on section 5.3.2, ΔU can then be expressed as:

2. It can be the case when skins are made of metal and the core is made of paper honeycombs or of viscoelastic foam.

$$\begin{aligned}
 \Delta U = -\pi \iint_S & \left[\eta_x^H D_1 \left(\frac{\partial \Phi_x}{\partial x} \right)^2 + \eta_y^H D_2 \left(\frac{\partial \Phi_x}{\partial x} \frac{\partial \Phi_y}{\partial y} \right) + \eta_y^H D_3 \left(\frac{\partial \Phi_y}{\partial y} \right)^2 + \dots \right. \\
 & \eta_{yz}^H D_4 \left(\Phi_y^2 - 2\Phi_y \frac{\partial w_0}{\partial y} + \left(\frac{\partial w_0}{\partial y} \right)^2 \right) + \eta_{xz}^H D_5 \left(\Phi_x^2 - 2\Phi_x \frac{\partial w_0}{\partial x} + \left(\frac{\partial w_0}{\partial x} \right)^2 \right) + \dots \\
 & \left. \eta_{xy}^H D_6 \left(\left(\frac{\partial \Phi_x}{\partial y} \right)^2 + 2 \frac{\partial \Phi_x}{\partial y} \frac{\partial \Phi_y}{\partial x} + \left(\frac{\partial \Phi_y}{\partial x} \right)^2 \right) \right] dx dy
 \end{aligned} \quad (5.16)$$

The kinetic energy T of the system is given in Eq. (5.17) as a function of Φ_x , Φ_y , and w_0 . In this expression, ρ^H is the density of the equivalent homogeneous thick plate. It is given by $h\rho^H = h^c\rho^c + 2h^s\rho^s$.

$$T = \frac{\rho^H \omega^2}{2} \iiint_{(V)} [u^2 + v^2 + w^2] d\tau = \frac{\rho^H \omega^2}{2} \iint_{(S)} \left[\frac{h^3}{12} (\Phi_x^2 + \Phi_y^2) + h w_0^2 \right] dx dy \quad (5.17)$$

5.3.3 Estimation of modal parameters by an extended Rayleigh-Ritz procedure

In order to compare experimental results to numerical simulations, it is necessary to evaluate the frequencies f_n^{Num} and damping factors α_n^{Num} of the numerical modes corresponding to the sandwich panel. The dynamics of the plate is given by the hypotheses listed in section 5.3.2, by Eqs. (5.8), and by the boundary conditions. Under the “light damping” hypothesis, which assumes that modal shapes and frequencies are unchanged by the addition of damping, the frequencies f_n^{Num} of the numerical modes are easily accessible. The problem consists thus in evaluating the relationships between the α_n^{Num} damping factors and the η^H loss-factors. Notations used in this section are summarized by Tab. B.4 in the Appendix B.4.

Light damping hypothesis

The honeycomb sandwich panel is considered here as a non-conservative system \mathcal{P}_{NC} having N degrees of freedom. The associated conservative system, without hysteretic damping and having also N degrees of freedom, is denoted \mathcal{P}_{C} . The N modes of \mathcal{P}_{C} have their modal shapes denoted by ξ_n^{C} and their real modal frequencies denoted by f_n^{C} . The N modes of \mathcal{P}_{NC} have their modal shapes denoted by ξ_n^{NC} and their complex modal frequencies denoted by f_n^{NC} .

If \mathcal{P}_{NC} is lightly damped, it can be shown [Geradin & Rixen 1997] that $\xi_n^{\text{NC}} \simeq \xi_n^{\text{C}}$ and that $f_n^{\text{NC}} \simeq f_n^{\text{C}} + j \frac{\alpha_n}{2\pi}$ to first order. The “light damping” hypothesis thus assumes that modal shapes and real parts of the frequencies are unchanged by the addition of damping. This hypothesis has been shown to be acceptable for values of material loss factors lower than 0.1 [Cortes & Elejabarrieta 2006]. This assumption is similar to the assumption made in the “Modal Strain Energy” approach used to model sandwich panels having visco-elastic cores [Johnson & Kienholz 1982, Cortes & Elejabarrieta 2006, Torvik & Runyon 2007].

Let U_n^{NC} be the potential energy associated with the n^{th} mode of \mathcal{P}_{NC} for a maximum vibrational amplitude of 1 on the plate. It varies in time as $\exp(-2\alpha_n t)$ so that the energy lost by this mode during one cycle, ΔU_n^{NC} , is:

$$\Delta U_n^{\text{NC}} = -2 \frac{\alpha_n}{f_n^{\text{C}}} U_n^{\text{NC}} \quad (5.18)$$

Since \mathcal{P}_C and \mathcal{P}_{NC} have the same modal shapes, *i.e.* $\xi_n^{NC} \simeq \xi_n^C$, and the potential energy depends only on the modal shapes (see Eq. (5.13)), then $U_n^{NC} = U_n^C$. And since for the conservative system \mathcal{P}_{NC} , the equality $U_n^C = T_n^C$ is true, one then obtains:

$$\Delta U_n^{NC} = -2 \frac{\alpha_n}{f_n^C} T_n^C \quad (5.19)$$

Thanks to the light damping hypothesis, Eq. (5.19) gives a straightforward way to obtain the modal dampings α_n from ΔU_n^{NC} , T_n^C and, f_n^C .

Derivation of f_n^{Num}

A Rayleigh-Ritz procedure has been used to derive numerically the modal frequencies f_n^{Num} and the mode shapes ξ_n^{Num} of \mathcal{P}_C . To this end, the generalised-displacement fields $\Phi_x(x, y)$, $\Phi_y(x, y)$, and $w_0(x, y)$ are projected on the elements of an orthonormal polynomial basis of order Q satisfying partially the free-free boundary conditions [Bhat 1985, Kim 1986]:

$$\Phi_x(x, y) = \sum_{i,j} L_{ij} p_i(x) p_j(y) \quad \Phi_y(x, y) = \sum_{i,j} M_{ij} p_i(x) p_j(y) \quad w_0(x, y) = \sum_{i,j} N_{ij} p_i(x) p_j(y) \quad (5.20)$$

where the polynomials $p_i(\kappa)$ are generated as described in Appendix B.1.

This procedure generates a new set of $3Q^2$ generalised displacements L_{ij} , M_{ij} and N_{ij} . The next step consists in writing the kinetic and potential energies T and U which have been expressed as functions of Φ_x , Φ_y , and w_0 in section 5.3.2. The Hamilton principle reads as:

$$\forall (i, j) \in [0, Q-1]^2 : \frac{\partial(T-U)}{\partial L_{ij}} = 0 \quad \frac{\partial(T-U)}{\partial M_{ij}} = 0 \quad \frac{\partial(T-U)}{\partial N_{ij}} = 0 \quad (5.21)$$

The above system of $3Q^2$ linear equations can be re-written as $[\mathbf{K} - 4\pi^2 f^2 \mathbf{M}] \mathbf{q} = 0$. The expressions of the partial derivatives of U with respect to L_{ij} , M_{ij} , N_{ij} yield \mathbf{K} while the partial derivatives of T with respect to L_{ij} , M_{ij} , N_{ij} yield \mathbf{M} . The explicit expressions for these partial derivatives as functions of the p_i -s and of the generalised displacements are given in Appendix B.2. The resolution of this eigenvalue problem gives a straightforward access to the modal frequencies f_n^{Num} and mode shapes ξ_n^{Num} of \mathcal{P}_C . Under the light damping hypothesis, ξ_n^{Num} and f_n^{Num} found for \mathcal{P}_C are also the modal shapes and frequencies of \mathcal{P}_{NC} .

Derivation of α_n^{Num}

Introducing the modal coefficients ξ_n^{Num} (expressed in the $\{L_{ij}, M_{ij}, N_{ij}\}$ system of coordinates) into Eqs. 5.20 yields analytical expressions for the Φ_x , Φ_y , and w_0 modal fields and also for their x - and y -derivatives.

For each of the N modes, the potential, lost, and kinetic energies can be written by introducing these expressions into Eqs. 5.13, 5.16, and 5.17:

$$\forall n \in [1, N] : \quad T_n^C = 4\pi^2 (f_n^{\text{Num}})^2 t_n \quad U_n^{\text{NC}} = U_n^C = \sum_{k=1}^6 D_k u_k^n \quad \Delta U_n^{\text{NC}} = -\pi \sum_{k=1}^6 \eta_k D_k u_k^n \quad (5.22)$$

where the subscripts $\{1, 2, 3, 4, 5, 6\}$ of η stand for $\{x, \nu, y, yz, xz, xy\}$ respectively. The coefficients t_n and u_k^n are given explicitly in Appendix B.3. They depend on the geometry and mass parameters of the plate, and are quadratic in modal shapes ξ_n^{Num} . Note that t_n can be considered as half the modal mass for some normalised displacement. The product $D_k u_k^n$ represents the k -contribution to the n -th modal stiffness (where k stands for x, ν, y, yz, xz , or xy).

The expression (5.23) of the modal dampings α_n^{Num} can be deduced from the Eqs. (5.19) and the last two expressions of (5.22) or, equivalently, by (5.18) and the first two expressions of (5.22):

$$\alpha_n^{\text{Num}} = -\frac{f_n^{\text{Num}} \Delta U_n^{\text{NC}}}{2T_n^C} = \frac{1}{8\pi f_n^{\text{Num}} t_n} \sum_{k=1}^6 \eta_k D_k u_k^n \quad (5.23a)$$

$$\text{or } \alpha_n^{\text{Num}} = -\frac{f_n^{\text{Num}} \Delta U_n^{\text{NC}}}{2U_n^C} = \frac{f_n^{\text{Num}} \pi}{6 \sum_{m=1}^6 D_m u_m^n} \sum_{k=1}^6 \eta_k D_k u_k^n \quad (5.23b)$$

One can notice that α_n^{Num} is a linear combination of the η_k . This set of equations is a generalisation to all the modes of a thick-plate of the expression given by De Visscher *et al.* [de Visscher et al. 1997, Eq. (13)] for three particular modes of a thin plate. This expression is also a generalization to all the loss factors of the expression established by Johnson *et al.* [Johnson & Kienholz 1982] for sandwich structures having a visco-elastic core.

5.3.4 Estimation of modal parameters by high-resolution modal analysis

In this section, the experimental protocol that has been followed to obtain impulse responses and to extract the modal frequencies f_n^{XP} and the damping factors α_n^{XP} is presented. This protocol combines the procedures and implementation details presented by Rébillat et al. [2010b, 2011c] and Ege et al. [2009].

Experimental setup

Throughout all measurements, panels were suspended by thin wires in order to ensure free-free boundary conditions. Light panels were acoustically excited by an electro-dynamical loudspeaker driven by a wide-band electrical signal [Matter et al. 2009]. The velocity response was measured in one corner of each panel with a laser Doppler vibrometer (Ometron VH300+ type 8329). Eventual non-linearities arising from the loudspeaker were removed and the impulse response of the panel was reconstructed [Rébillat et al. 2010b, 2011c]. Since heavy panels can hardly be excited by acoustical means, impact excitations were also used; in this case, the acceleration resulting from the impact was recorded with a light accelerometer fixed in the vicinity of one corner of the panel. The impulse response was obtained after deconvolution with the

nearly-impulsive force signal [Ege et al. 2009]. Since no nodal line goes through corners of a free vibrating plate, all excited modes contribute to the resulting impulse response.

High resolution modal analysis

It is hypothesised that the experimental data are corrupted by additive noise. Thus, an impulse response $h(t)$ is mathematically represented as a sum of decaying exponentials (natural modes) and measurement noise $b(t)$ as in Eq. (5.24). Each contribution of a natural mode is characterised by its amplitude A_n^{XP} , frequency f_n^{XP} , damping α_n^{XP} and phase ϕ_n^{XP} .

$$\begin{aligned} h(t) &= \sum_{n=1}^N A_n^{\text{XP}} \exp(-\alpha_n^{\text{XP}} t) \cos(2\pi f_n^{\text{XP}} t + \phi_n^{\text{XP}}) + b(t) \\ &= \sum_{n=1}^N \frac{A_n^{\text{XP}}}{2} [\exp(-\alpha_n^{\text{XP}} t + j2\pi f_n^{\text{XP}} t + j\phi_n^{\text{XP}}) + \exp(-\alpha_n^{\text{XP}} t - j2\pi f_n^{\text{XP}} t - j\phi_n^{\text{XP}})] + b(t) \end{aligned} \quad (5.24)$$

In order to extract experimental modal frequencies f_n^{XP} and damping α_n^{XP} from $h(t)$, the recently developed "High Resolution Modal Analysis" (HRMA) [Ege et al. 2009] has been applied and is briefly sketched below. The signal is projected onto two subspaces: the subspace spanned by the sinusoids (*signal subspace*) and its supplementary (*noise subspace*) according to the ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) algorithm [Roy & Kailath 1989]. The frequencies f_n^{XP} and dampings α_n^{XP} of a given number of modes are the eigenvalues of a matrix obtained after some computation on the observed signal. The amplitudes A_n^{XP} and phases ϕ_n^{XP} are estimated afterwards by a least-mean-square method.

In the ESPRIT procedure, the dimensions of both subspaces must be chosen *a priori* and the quality of the estimation significantly relies on a proper choice for these parameters. The best choice for the dimension of the signal subspace is the number of exponentials (twice the number of decaying sinusoids, or real modes, see Eq. (5.24)). This number can be estimated, before the analysis, by means of the ESTER technique [Badeau et al. 2006].

To improve the performance of the ESPRIT algorithm, signals are split into several frequency-bands [Laroche 1993], thus reducing the number of modes to be processed. In order to limit computation time, the responses of the band-pass filters are frequency-shifted and down-sampled.

A block-diagram describing the different steps involved in HRMA is shown in Fig. 5.9.

Uncertainties in modal parameters estimation

To give an overview of the precision offered by the HRMA, this method is applied to a synthetic signal obtained by adding two decaying exponentials of equal amplitudes to white noise. The sampling frequency is $f_s = 44.1$ kHz. The two modal frequencies are 592 and 596 Hz, very close one from each other. For relatively important modal dampings, these two modes overlap in the frequency-domain and therefore, their frequencies and dampings are difficult to estimate with methods based on the Fourier-Transform (FT). The modal overlap factor μ (*i.e.* the ratio between the half-power modal bandwidth $\Delta f_{-3\text{ dB}}$ and the average

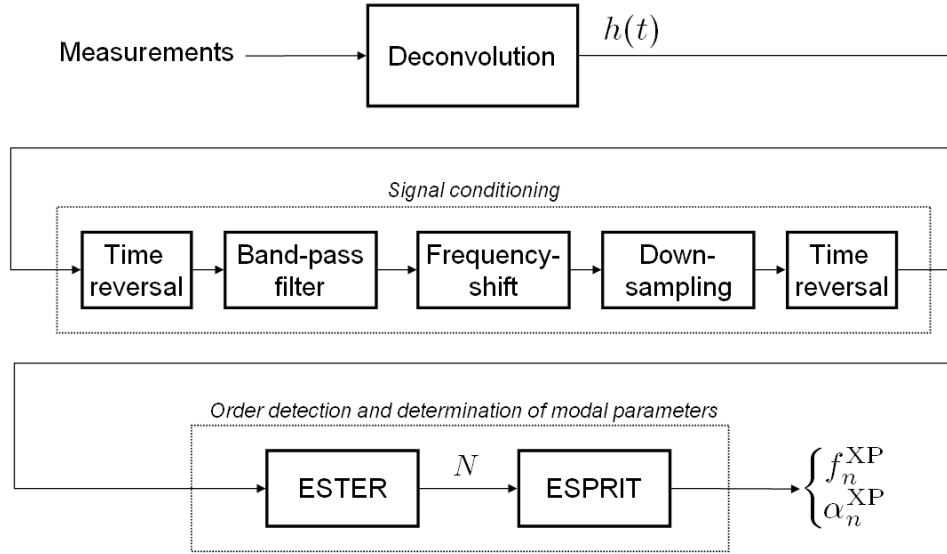


Figure 5.9: Block diagram of the high resolution modal analysis method, adapted from [Ege et al. 2009]. The deconvolution block is described in [Rébillat & Boutillon 2010a].

modal spacing Δf_{mode}) quantifies this phenomenon [Ege et al. 2009]. If the modal damping α is the same for both modes, μ is:

$$\mu = \frac{1}{\Delta f_{\text{mode}}} \frac{\alpha}{\pi} \quad (5.25)$$

In practice, the FT cannot efficiently separate modes when $\mu > 30\%$ [Berthaut et al. 2003]. For HRMA, this upper limit depends on the signal to noise ratio and on the number of components which are retained in the pre-conditionning step. As an example, estimations of modal frequencies and dampings have been performed on the synthetic signal with various α -values (corresponding to modal overlaps from $\mu = 1\%$ to 150%) and a signal-to-noise ratio (SNR) increasing from 10 dB to 50 dB. SNR is understood here as $20 \log_{10} \left(\frac{S_{RMS}}{B_{RMS}} \right)$, where S_{RMS} is the root mean square (RMS) value of signal in absence of noise and B_{RMS} the RMS value of noise in absence of signal. For each couple $\{\mu, \text{SNR}\}$, modal parameters were estimated 50 times. For each mode of the synthetic signal, the uncertainty in frequency or damping is defined as the mean of the absolute values of the relative error between the original and the estimated data. These uncertainties, expressed in dB, are shown as contour plots in Fig. 5.10.

It can be seen in Fig. 5.10 that the uncertainties are very small for both dampings and frequencies even for high modal overlaps and low SNR. In general, the uncertainty increases with μ and decreases slightly when SNR increases. The HRMA gives better estimations of the eigenfrequencies and dampings than the FFT for a wide range of μ and SNR. Since the modal overlap μ generally increases in the impulse response of a plate, and the SNR decreases with the frequency, the HRMA gives access to significantly more modes than the usual FT-based modal-analysis techniques and is thus of great interest in the present context.

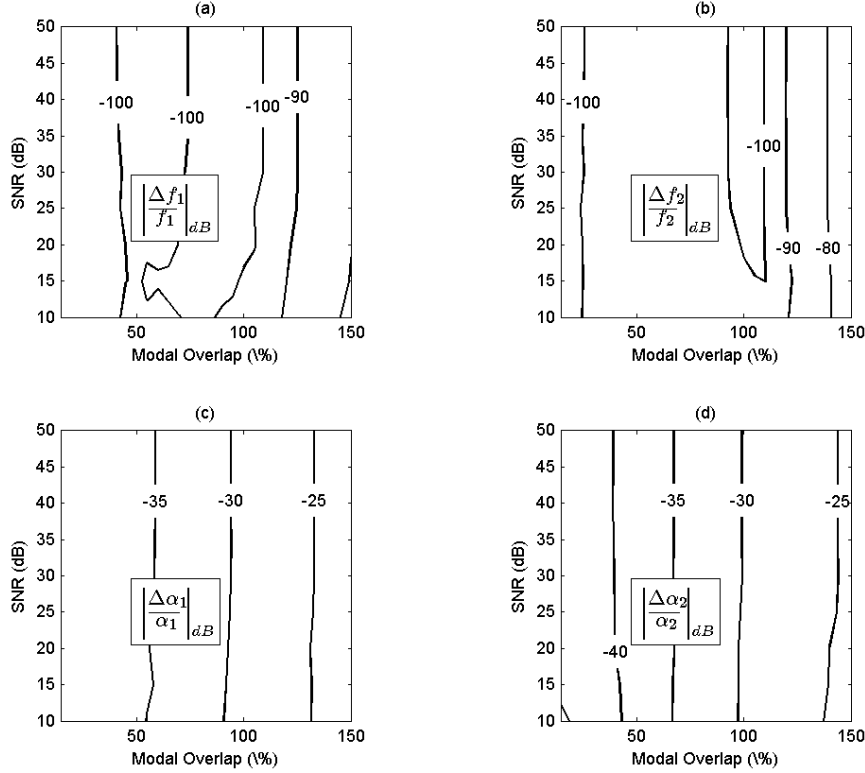


Figure 5.10: Contour plots of the uncertainty on the estimated "modal" frequencies and dampings of a synthetic signal containing two decaying sinusoids as a function of the modal overlap and the signal-to-noise ratio. The uncertainty is the mean of 30 absolute values of the relative error between original and estimated data, for 30 different realisations of the noise. (a): uncertainties on the frequency of the first mode. (b): uncertainties on the frequency of the second mode. (c) and (d): *idem* relatively to the dampings.

5.3.5 Optimisation procedure

This section describes how to derive, in two steps, the complex moduli of elasticity of the homogenised equivalent material of the sandwich plate $\{E_x^H, \eta_x^H, E_y^H, \eta_y^H, G_{xy}^H, \eta_{xy}^H, G_{xz}^H, \eta_{xz}^H, G_{yz}^H, \eta_{yz}^H, \nu_{xy}^H, \eta_\nu^H\}$ from the experimental and numerical values of the modal frequencies and dampings f_n^{XP} , f_n^{Num} , α_n^{XP} , and α_n^{Num} .

Elastic properties

The estimation of the elasticity parameters $\{E_x^H, E_y^H, G_{xy}^H, G_{xz}^H, G_{yz}^H, \nu_{xy}^H\}$ is done by comparing the experimental and numerical modal frequencies. The estimation problem to solve is non-linear and several orders of magnitude are involved in the properties values. The following cost-function was used:

$$C_f = \sum_{n=1}^N \left(\frac{f_n^{XP} - f_n^{Num}}{f_n^{XP}} \right)^2 \quad (5.26)$$

A steepest-descent (with backtracking) algorithm [Kelley 1999] using rigidities $\{D_1, D_2, D_3, D_4, D_5, D_6\}$ as design variables has been chosen. In the present case, the coefficients

of the gradient can be easily derived analytically, making the method easy to implement and computationally light.

Estimation results obtained by gradient methods are known to be very dependent on the initial values of the parameters. To minimise the influence of the starting point, the following initialisation strategy for the rigidities has been chosen³ :

1. Initial values of in-plane rigidities D_1 , D_2 , D_3 and D_6 are the most influential; they were derived from the three lowest modal frequencies of the panel, as proposed in [Mac Intire & Woodhouse 1988].
2. Initial values of out-of-plane rigidities D_4 and D_5 are less critical; homogenisation theory proposed by Gibson [Gibson & Ashby 2001] for honeycomb core sandwich panels is used. This theory requires a value for the elasticity moduli of the material composing the honeycomb core. The first estimation was based on static tests.

In the following examples, 10 iterations were enough to reach convergence: 10^{-7} for the gradient.

Damping properties

As can be seen in Eq. (5.23), modal dampings depend linearly on the loss factors $\{\eta_x^H, \eta_y^H, \eta_{xy}^H, \eta_{xz}^H, \eta_{yz}^H, \eta_\nu^H\}$ once the rigidities have been found. The estimation of the loss factors is therefore much easier than that of the elasticity parameters. A simple least-square optimisation procedure is sufficient to estimate the loss factors from the modal dampings.

The following cost-function has been chosen:

$$C_\alpha = \sum_{n=1}^N \left(\frac{\alpha_n^{XP} - \alpha_n^{Num}}{\alpha_n^{XP}} \right)^2 \quad (5.27)$$

The optimisation procedure is not iterative and needs no particular initialisation.

Determination of the order Q of the polynomial basis and of the number N of included modal parameters

Two parameters have to be chosen in order to apply the optimisation procedures described in Sec. 5.3.5. These methodological parameters are the order Q of the polynomial basis associated with the extended Rayleigh-Ritz procedure (see Sec. 5.3.3) and the number N of modal parameters (frequencies or dampings) to be included in the optimisation procedure.

The parameters Q and N can be chosen differently for the estimation of the elastic material properties and for the estimation of the material loss factors respectively. For the estimation of the elastic material properties, the parameters Q_f and N_f that minimize the cost functions C_f given by Eq. (5.26) are chosen. For the estimation of the damping properties, the parameters Q_α and N_α that minimize the cost functions

3. Il n'y a pas, a priori, unicité de la solution de la procédure d'optimisation. Il est en effet possible que différents jeux de paramètres d'élasticité conduisent à un même jeu de fréquences modales. Pour éviter cette écueil, des paramètres d'élasticité ayant un réel sens physique ont été cherchés. La procédure d'initialisation décrite par Mac Intire & Woodhouse [1988] définit un point de départ possédant un sens physique et donc permet dans une certaine mesure d'éviter de converger vers des solutions non-physiques.

C_α given by Eq. (5.27) are chosen. Selecting the optimal Q and N values is done empirically by running the optimisations for different values of these parameters, typically in the ranges $N \in \{20 : 40\}$ and $Q \in \{10 : 18\}$.

Sensitivity analysis

For the estimation procedures described in section 5.3.5 to be efficient, modal frequencies and dampings must convey a sufficient amount of information relative to each parameter to be estimated. In other words, modal dampings and frequencies have to be sensitive to the parameters of interest.

Relevant sensitivities can be defined and calculated analytically. The sensitivity of the modal value τ to the parameter γ is noted S_γ^τ and defined by Eq.(5.28): if γ is increased by 1%, τ increases by $S_\gamma^\tau\%$.

$$S_\gamma^\tau = \frac{\Delta\tau}{\Delta\gamma} \frac{\gamma}{\tau} = \frac{\frac{\Delta\tau}{\tau}}{\frac{\Delta\gamma}{\gamma}} \quad (5.28)$$

According to Eq. (5.22) and with the same notations, the sensitivity $S_{D_k}^{f_n}$ of the n^{th} modal frequency f_n to the rigidity $D_k \in \{D_1, D_2, D_3, D_4, D_5, D_6\}$ can be written as:

$$S_{D_k}^{f_n} = \frac{D_k u_{nk}}{8\pi^2 f_n^2 t_n} \quad (5.29)$$

Similarly, using Eq. (5.23), the sensitivity $S_{\eta_k}^{\alpha_n}$ of the n^{th} modal damping α_n to the loss factor $\eta_k \in \{\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6\}$ can be written as:

$$S_{\eta_k}^{\alpha_n} = \frac{\eta_k D_k u_{nk}}{4\pi f_n \alpha_n t_n} \quad (5.30)$$

The amount of information relative to one given parameter and contained in one given mode can be easily quantified with Eq. (5.29) and Eq. (5.30). Examples are given in Figs. 5.13, 5.16, and 5.18 (see Sec. 5.3.7).

5.3.6 Validation of the estimation procedure

A validation of the mechanical model and procedures given in sections 5.3.2, 5.3.3, and 5.3.5 is proposed as follows: the experimental results of the modal analysis are replaced by those of the simulation of a finite-element (FEM) of two virtual plates with known properties. The modes of a homogeneous thick plate as modelled using the FEM are first compared to those given by the extended Rayleigh-Ritz procedure applied to the mechanical model given in section 5.3.2. The method for deriving elasticity and damping parameters as sketched in Fig. 5.7 is afterwards validated on a virtual sandwich plate.

Finite-element model of the virtual plates

To validate the extended Rayleigh-Ritz procedure applied to the mechanical model, a homogeneous thick virtual plate was designed. A FE-model of the sandwich panel has also been built to test the accuracy of the estimation method. The chosen sandwich plate is made of 3 homogeneous layers and is symmetrical

	l_x (m)	l_y (m)	h (m)		ρ (kg/m ³)	
HVP	0.4	0.6	4×10^{-3}		700	
VSP	0.4	0.6	Core	Skin	Core	Skin
			4×10^{-3}	0.2×10^{-3}	40	700

Table 5.1: Geometry and constituent densities of the homogeneous virtual plate (HVP) and of the virtual sandwich-plate (VSP).

		E_x	E_y	ν_{xy}	G_{xy}	G_{xz}	G_{yz}
HVP	Real part	4 GPa	5 GPa	0.33	1 GPa	10^{-2} GPa	10^{-2} GPa
	Loss factor (%)	2	5	1	0.5	1	1
VSP Skins	Real part	4 GPa	5 GPa	0.33	1 GPa	1 GPa	1 GPa
	Loss factor (%)	2	5	1	0.5	1	1
VSP Core	Real part	1×10^{-3} GPa	1×10^{-3} GPa	0.33	1×10^{-3} GPa	1×10^{-2} GPa	3×10^{-2} GPa
	Loss factor (%)	1	1	1	1	3	5

Table 5.2: Mechanical parameters chosen for the the homogeneous virtual plate (HVP) and for the virtual sandwich-plate (VSP)

with respect to its mid-plane. Geometrical, mechanical and mass parameters of the two plates are given in Tab. 5.1 and Tab. 5.2.

For the two virtual plates under study, the finite element model is built on a 2D rectangular mesh made of 60 by 60 regularly spaced points. This value is justified at the end of this section. At each point an 8-nodes shell element is placed with a linear expansion of the in-plane displacements in the thickness coordinate and a constant transverse displacement through the thickness (COQ8 of the Cast3M code [Verpeaux et al. 1988]). Each of these elements possess 6 degrees of freedom (the translations in the x, y, and z directions and the rotations around the x-, y- and z-axes). In the case of the virtual sandwich plate, the three-layers are modelled as one equivalent layer as in Sec. 5.3.2.

Modal frequencies and dampings of the virtual plates

Finite-element modelling and the associated computations have been performed using Cast3M [Verpeaux et al. 1988], a free software developed by the French Centre for Atomic Energy (CEA). This software is used here as it allows to find the complex modes (modal shapes ξ_n^{FEM} , dampings α_n^{FEM} and frequencies f_n^{FEM}) of a problem put in the form: $\mathbf{K} + j\omega\mathbf{C} - \omega^2\mathbf{M} = \mathbf{0}$. In this formulation, \mathbf{M}, \mathbf{C} and \mathbf{K} must be real matrices to be accepted by the VIBC function of the Cast3M code. Complex modes are then found by solving a complex valued generalized eigenvalue problem using a QZ-algorithm. According to the possibilities offered by Cast3M, modal frequencies and dampings of the virtual plates are computed using the following procedure:

1. The conservative system is described according to the constitutive model of section 5.3.2 with no hysteretic damping taken into account. A mass matrix \mathbf{M} and a real stiffness matrix \mathbf{K}' are deduced from this model.
2. The N first modal frequencies $\{f_n^{\text{FEM}}\}_{n \in [1, N]}$ of the conservative system are computed by solving, in the Fourier domain, the real-valued eigenvalue problem $\mathbf{K}' - \omega^2\mathbf{M} = \mathbf{0}$.

3. The non-conservative system is described according to the constitutive model of section 5.3.2, including hysteretic damping. A mass matrix M and a complex stiffness matrix $K = K' + jK''$ are deduced from this model.
4. The "light damping hypothesis," is retained. The real part of the modal frequencies of the non-conservative system are thus already known (see Sec. 5.3.3).
5. For each one of the N first modes of the non-conservative system, the following operations are then performed:
 - (a) The dynamic equation of the dissipative system are formulated, in the Fourier domain, as $-\omega^2 M q + j\omega C q + K' q = 0$ with $C = K' / (2\pi f_n^{\text{FEM}})$. The problem is thus formulated as expected by the VIBC function with C real but frequency-dependent. Its important to notice that this equation models correctly the hysteretic damping model described in Sec. 5.3.2 only near $\omega \simeq 2\pi f_n^{\text{FEM}}$.
 - (b) The modal loss factor α_n is obtained as the imaginary part of the eigenvalue of this new problem solved near $\omega \simeq 2\pi f_n^{\text{FEM}}$.

For the homogeneous virtual plate, increasing the number of elements above 60 elements per side results in less than a 1% relative variation of the 35 first modal frequencies (conservative and non conservative cases) and in less than 0.4% of the 35 first modal dampings. The same convergence is observed for the 3-layer virtual sandwich plate. Thus, 60 elements per side are enough to ensure the desired precision on the analysis of the first 35 modes of the two plates.

Validation of the extended Rayleigh-Ritz resolution of the mechanical model

Comparing the modal frequencies and dampings given by the extended Rayleigh-Ritz resolution method (18-order) and by the FE-model for the homogeneous thick virtual plate provides an estimate of the reliability of the retained mechanical model coupled with the extended Rayleigh-Ritz resolution method for the first 35 modes.

The plate under study in the present section is not a sandwich panel as the one modelled in Sec. 5.3.2 but a homogeneous thick plate. For such homogeneous plates, values for the shear correction factors are usually chosen between $2/3$ and $5/6$ instead of 1, which is recommended only for sandwich panels [Birman 2002]. Values of $\kappa_{yz} = \kappa_{xz} = 0.7$ have been arbitrarily chosen here in the range $\{0.666 : 0.833\}$.

With shear correction factors $\kappa_{yz} = \kappa_{xz} = 0.7$, the mean absolute difference between FE-results and Rayleigh-Ritz results is lower than 1.2% for modal frequencies and lower than 25% for modal dampings. The larger error on modal dampings may be explained as follows. The damping matrix C provided to the FEM results from the writing of the stiffness matrix and is not necessarily diagonal when expressed in the basis of the conservative modes. By contrast, the extended Rayleigh-Ritz procedure accounts for dissipation by associating one damping coefficient α_n^{Num} to each mode, neglecting by construction any coupling between conservative modes.

Estimation results for the 3-layer virtual sandwich-plate

A 3-layer virtual sandwich-plate has been used to validate the estimation procedure described in Fig. 5.7. Since, for this 3-layer virtual sandwich-plate, convergence problems were encountered with the steepest-

descent algorithm, a simplex search method [Lagarias et al. 1998] was used instead in this case (function “*fminsearch*” in Matlab™). The initialisation procedure remains the same as the one described in Sec. 5.3.5. For all other optimisations, the steepest-descent algorithm is used.

Based on the first $N = 35$ modal frequencies given by the FEM and using a Rayleigh-Ritz order $Q = 16$, the estimated values of the elasticity parameters are compared to the original values given to the FEM. The loss factors have been estimated with $N = 28$ modes and a model order $Q = 18$. The estimated mechanical parameters are presented in Fig. 5.11 for each layer of the sandwich.

The residual mismatch between the results of estimation and the original values is discussed here. The mean absolute value $\left\langle \left| \frac{\Delta f_n}{f_n} \right| \right\rangle$ of the relative difference between experimental and numerical modal frequencies is 2.6%. For the dampings, the residual mismatch $\left\langle \left| \frac{\Delta \alpha_n}{\alpha_n} \right| \right\rangle$ is 21.6%. These orders of magnitude, as compared to the one obtained in Sec. 5.3.6, suggest that the assumption that a 3-layer sandwich plate can be modelled as a simple homogeneous thick plate is correct in the frequency range under study.

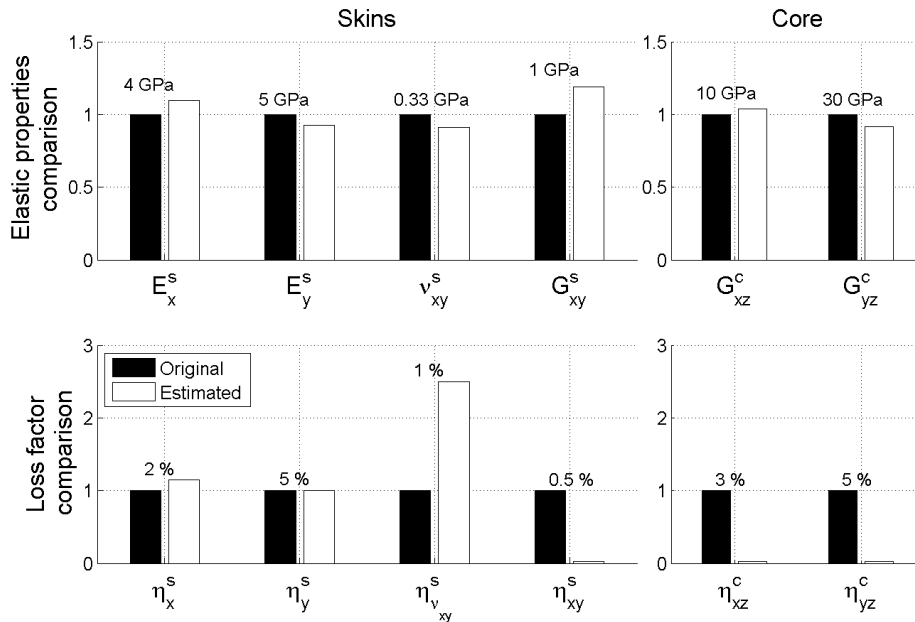


Figure 5.11: Comparison between the values of the mechanical parameters used in the FEM (Original), and their estimated values (Estimated) for each constituent (Skins, Core) of the virtual sandwich plate. Original parameters are represented as black bars with their numerical value indicated above. White bars represent the ratio of the estimated to the original parameters. Elasticity parameters have been estimated with 35 modes and with a model order $Q = 16$. Loss-factors have been estimated with 28 modes and with a model order $Q = 18$.

It can be seen in Fig. 5.11 that the agreement between estimated and original parameters is globally very good. In-plane elasticity parameters of the skins and out-of-plane elasticity parameters of the core are estimated with a mean absolute relative error of 10.2%. Principal in-plane loss-factors η_x^s and η_y^s are estimated with a comparable accuracy of 7.5%. The imaginary part of ν_{xy}^s is largely overestimated while the imaginary part of G_{xy}^s is underestimated. However, the overestimation of one parameter may be the result of the underestimation of the other, by compensation. The imaginary parts of G_{xz}^c and G_{yz}^c

are assigned zero values by the estimation process. The fact that zeros values are found illustrates the limitations of the thick-plate model under Reissner-Minldin hypothesis. Physically, this underestimation is due to the fact that only a marginal part of the total energy-loss per cycle is dissipated through the mechanical couplings described by \underline{G}_{xz}^c and \underline{G}_{yz}^c . Modal dampings factors are thus here not very sensitive to these material loss factors.

A complete validation study should have established the validity limits of the estimation method. Even though this is not what has been done here, the above results suggest that the mixed numerical/experimental procedure is potentially an accurate tool for the estimation of the main elasticity moduli and loss-factors of 3-layers sandwich plates.

5.3.7 Experimental results

Plate specimens

Three different sandwich panels with heterogeneous cores have been investigated. The first two – HC₁ and HC₂ – are rectangular lightweight honeycomb sandwich panels. Their skins and honeycomb cores are made of epoxy and paper. The third panel, made of two stainless steel sheets (skins) and two bidirectionally corrugated steel layers with a 20% relative density (core), is denoted CC (for "corrugated core"). The geometry and mass parameters of each panel are given in Tab. 5.3.

	l_x (cm)	l_y (cm)	h^s (mm)	h^c (mm)	s_{cell} (mm)	ρ^c (kg/m ³)	ρ^s (kg/m ³)
HC ₁	39.15	59.10	0.20	4.88	4.0	37.8	713
HC ₂	80.00	99.95	0.20	3.80	4.0	37.8	713
CC	17.78	22.86	0.20	1.48	1.0	2164	7800

Table 5.3: Geometry and constituent-densities of three sandwich panels HC₁, HC₂, and CC. The characteristic side-length of the core-cells is s_{cell} .

The $\frac{E_{x,y}^s}{G_{yz,xz}^c} \simeq 100$ criterium (see Sec. 5.3.2) that must be satisfied turns out to be met for all the plates that have been tested. According to section 5.3.1, the sandwich core can be considered as homogeneous in the in-plane directions up to a given frequency f_{max} if the corresponding wavelength λ^{min} contains at least 50 cells [Burton & Noor 1997]. For a typical cell side-length s_{cell} and height h , this implies that $\frac{\lambda^{\text{min}}}{s_{\text{cell}}} > 50$. Moreover, plates must be thick-enough in order that out-of-plane elasticity parameters and loss-factors be estimated, but not too thick for the thick-plate theory to remain valid. This implies $\frac{2h}{\lambda^{\text{min}}} < 0.08$ [Liu & Zhao 2006]. The validity of these assumptions will be discussed.

Results for panel HC₁

Panel HC₁ was acoustically excited [Matter et al. 2009] and 46 modes were identified. Elasticity parameters and loss-factors were estimated with $N = 40$ modes and a model order $Q = 14$ using the steepest-descent algorithm of Sec. 5.3.5. The estimated parameters of the equivalent homogeneous plate and the corresponding skin and core parameters are given in Tab. 5.4.

		\underline{E}_x	\underline{E}_y	$\underline{\nu}_{xy}$	\underline{G}_{xy}	\underline{G}_{xz}	\underline{G}_{yz}
Equivalent plate	Real part	1.0 GPa	1.4 GPa	0.25	0.46 GPa	12 MPa	26 MPa
	Loss factor (%)	1.5	1.3	0	1.2	5.5	4.1
Core	Real part	-	-	-	-	12 MPa	26 MPa
	Loss factor (%)	-	-	-	-	5.5	4.1
Skins	Real part	4.8 GPa	6.8 GPa	0.25	5.6 GPa	-	-
	Loss factor (%)	1.5	1.3	0	1.2	-	-
Skins	Tensile tests	5.3 GPa	7.3 GPa	0.28	-	-	-
		± 0.5 GPa	± 0.7 GPa	± 0.04			

Table 5.4: Estimated mechanical parameters for panel HC₁. Parameters relative to the skins and the core are obtained after inversion of Eqs. (5.9) and (5.10).

The equivalent plate corresponding to panel HC₁ was found to be slightly orthotropic. This is a consequence of the laminated skins and of the orthotropy of the honeycomb structure. One can also notice that very low values are found for the loss factors associated with the Poisson ratio: little energy is dissipated via the Poisson effect in panel HC₁. Also, out-of-plane loss factors are relatively larger than the in-plane loss factors; this denotes that dissipation in honeycomb core structures is larger for out-of-plane shearing than for bending.

Static tensile tests have been performed on two samples of the skin material in the x- and y-directions respectively. The results are: $E_x^s = 5.3 \pm 0.5$ GPa, $\nu_{xy}^s = 0.28 \pm 0.04$, $E_y^s = 7.3 \pm 0.7$ GPa, $\nu_{yx}^s = 0.27 \pm 0.04$. These values are in excellent agreement with the values estimated using the proposed method. This constitutes an additional validation for the proposed method.

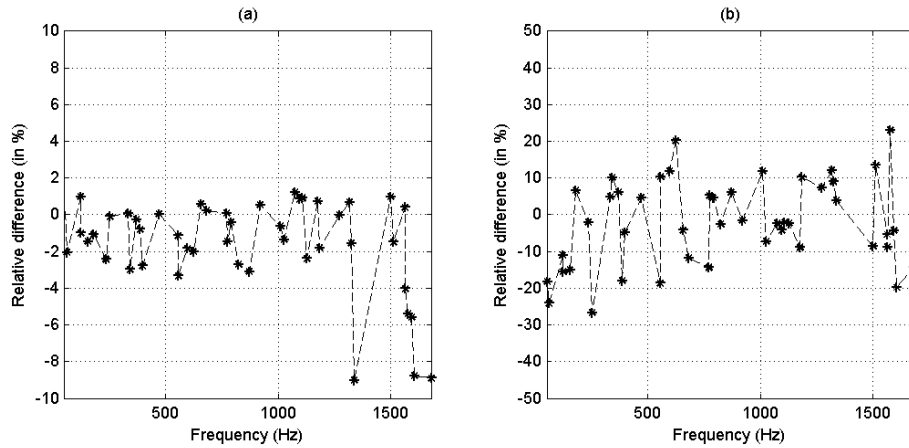


Figure 5.12: Residual differences on eigenfrequencies (a) and dampings (b) for panel HC₁. Numerical values have been obtained using elasticity and damping parameters from Tab. 5.4.

The optimisation procedure consists in minimising the difference between the experimental modal frequencies and dampings and numerical modal frequencies and dampings. The residual differences are presented in Fig. 5.12 and provide an estimation of the reliability of the method. One can observe that there is a very low relative difference between the measured and numerical modal frequencies: $\left\langle \left| \frac{\Delta f_n}{f_n} \right| \right\rangle = 2\%$.

Thus, the homogeneous thick plate model based on the Reissner-Mindlin hypothesis agrees with the real dynamical behavior of panel HC₁. Moreover, there is also a low relative difference between measured and numerical modal dampings: $\left\langle \left| \frac{\Delta \alpha_n}{\alpha_n} \right| \right\rangle = 10\%$. The frequency-independant loss factors combined with the "light damping" hypothesis appears to be a good model for the constitutive material of panel HC₁.

Based on the estimated values given in Tab. 5.4, the modal shapes can be computed by means of the extended Rayleigh-Ritz procedure. In the y -direction, panel HC₁ has a maximum of 8 nodal lines in the frequency range under consideration. This corresponds to $4.5\lambda^{\min}$: the shortest wavelength is 13 cm and contains $\lambda_y^{\min}/s_{\text{cell}} \simeq 30$ cells. In the x -direction, there are up to 6 nodal lines: $\lambda_x^{\min} = 11$ cm, $\lambda_x^{\min}/s_{\text{cell}} \simeq 28$ cells. Theoretically, this is hardly sufficient for the core to be considered as homogeneous. The condition $\frac{2h}{\lambda^{\min}} \leq 0.08$ is met for almost all modes since $\frac{2h}{\lambda^{\min}} = 0.08$ in the y -direction and $\frac{2h}{\lambda^{\min}} = 0.1$ in the x -direction. However, a very good agreement is observed in Fig. 5.12 between the homogeneous model and the experimental values. Therefore, 30 cells per wavelength may be enough in the present case for the core to be considered as homogeneous. The thick-plate theory also seems sufficient for $\frac{2h}{\lambda^{\min}} = 0.1$.

Sensitivities of the modal frequencies and dampings to the in-plane and out-of-plane mechanical parameters for panel HC₁ are shown in Fig. 5.13. Modal frequencies and dampings are sensitive to all the in-plane elasticity and damping parameters. The estimated in-plane mechanical properties are thus reliable. Sensitivities to the out-of-plane mechanical properties are relatively important. This ensure a high degree of confidence for the estimated values of $\text{Re}(\underline{G}_{xz}^H)$ and $\text{Re}(\underline{G}_{yz}^H)$ since a large number of sensitive modes are involved in the optimisation procedure. Sensitivity to the out-of-plane loss factors is one order of magnitude lower. Thus, estimated out-of-plane loss factors are less reliable than in-plane loss factors.

Results for panel HC₂

Panel HC₂ was excited by an impact hammer and 26 modes were extracted. Elasticity parameters and loss-factors were estimated with $N = 26$ modes and a model order $Q = 14$ using the steepest-descent algorithm of Sec. 5.3.5. The estimated parameters of the equivalent homogeneous plate are given in Tab. 5.5.

		\underline{E}_x	\underline{E}_y	ν_{xy}	\underline{G}_{xy}	\underline{G}_{xz}	\underline{G}_{yz}
Equivalent plate	Real part	1.0 GPa	1.2 GPa	0.27	0.48 GPa	13 MPa	28 MPa
	Loss factor (%)	1.0	1.1	0.0	3.2	44	30
Skins	Real part	3.8 GPa	4.7 GPa	0.27	1.9 GPa	-	-
	Loss factor (%)	1.0	1.1	0.0	3.2	-	-
Core	Real part	-	-	-	-	13 MPa	28 MPa
	Loss factor (%)	-	-	-	-	44	30

Table 5.5: Estimated mechanical parameters for panel HC₂. Parameters relative the skins and the core are obtained after inversion of Eqs. (5.9) and (5.10)

As can be seen in Fig. 5.14, in-plane and out-of plane elasticity and damping parameters are similar to those of panel HC₁. Theoretically, if the cores of the two plates were made of the same material (which is not known with certainty but seems to be the case), $\text{Re}(\underline{G}_{xz}^H)$ and $\text{Re}(\underline{G}_{yz}^H)$ should be equal for both panels, according to Eq. (5.9). This is verified here with a good degree of precision. This robustness against the

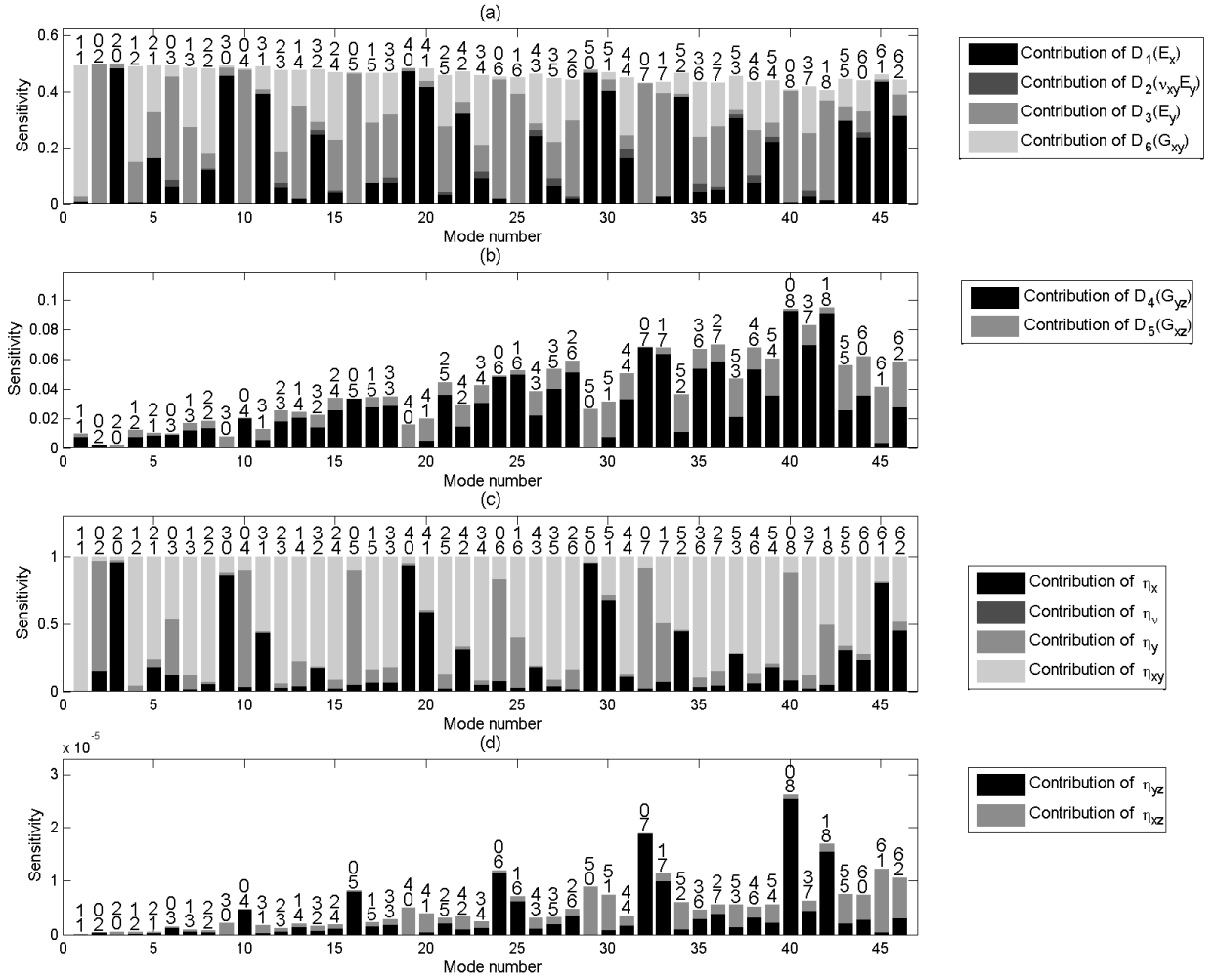


Figure 5.13: Sensitivities of the modal frequencies to the in-plane (a) and the out-of-plane (b) elasticity parameters. Sensitivities of the modal dampings to the in-plane (c) and the out-of-plane (d) loss factors for panel HC₁. Modes are identified by the numbers of their nodal lines in the x- and y-directions (top and bottom numbers respectively, on top of each bar).

size of test-panels constitutes an additional indication that the proposed method is reliable with regard to material properties.

The residual differences $\left\langle \left| \frac{\Delta f_n}{f_n} \right| \right\rangle$ and $\left\langle \left| \frac{\Delta \alpha_n}{\alpha_n} \right| \right\rangle$ for panel HC₂ are shown in Fig. 5.15. They are approximately 3.8% and 16.7% respectively. Plate and material models can be considered as appropriate for these honeycomb core sandwich panels (panels HC₁ and HC₂).

Based on the estimated values given in Tab. 5.5, the modal shapes of the extracted modes can be computed with the extended Rayleigh-Ritz procedure (Sec. 5.3.3). In the y -direction, panel HC₂ has a maximum of 5 nodal lines in the frequency range under consideration: its side-length is $3\lambda_y^{\min}$ and $\lambda_y^{\min} = 33$ cm, containing $\simeq 80$ cells. In the x -direction, panel HC₂ has a maximum of 5 nodal lines: $\lambda_x^{\min} = 26.5$ cm, $\frac{\lambda_x^{\min}}{s_{\text{cell}}} \simeq 65$ cells. This is more than sufficient for the core to be considered as homogeneous. Moreover in both directions, $\frac{2h}{\lambda_{\min}} \simeq 2.4 \times 10^{-2} < 0.08$, which is theoretically sufficient for modal frequencies to be

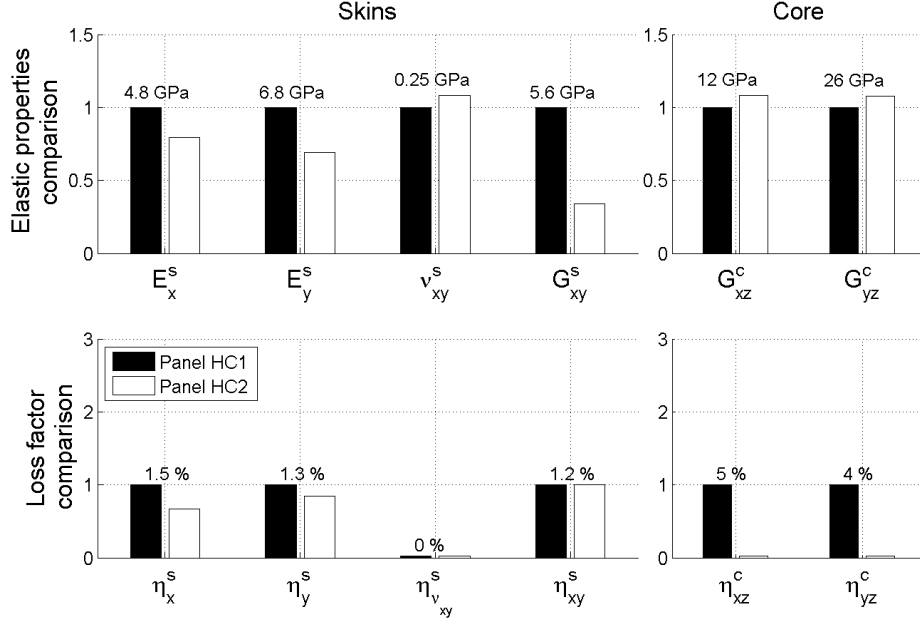


Figure 5.14: Comparison between the skins and core mechanical parameters estimated from panels HC₁ and HC₂. Numerical values indicated as references black bars correspond to the results obtained for the panel HC₁. White bars represent the ratio of the estimated value for HC₂ relatively to the one estimated for HC₁.

predicted using thick-plate theory. The core of panel HC₂ can be considered as homogeneous and all the modal frequencies can be predicted using thick-plate theory.

Sensitivities of the modal frequencies and dampings to the in-plane and out-of-plane mechanical parameters for panel HC₂ are shown in Fig. 5.16. Sensitivities to the elastic out-of-plane mechanical properties are lower for panel HC₂ than for panel HC₁. However, since results obtained on panel HC₂ are in close agreement with those obtained with panel HC₁, this suggests that a sensitivity of $\simeq 10^{-2}$ may still yield reliable results. Very low sensitivities to the out-of-plane loss factors explain that values of loss factors estimated on panel HC₂ deviate significantly from the values obtained with panel HC₁.

Results for panel CC

Panel CC was excited with an impact hammer. Elasticity material parameters have been estimated with $N = 35$ modes and a model order $Q = 14$ using the steepest-descent algorithm of Sec. 5.3.5. Loss factors have been estimated with $N = 23$ modes and a model order $Q = 13$. The estimated parameters of the equivalent homogeneous plate are given in Tab. 5.6.

At first, it can be seen from Tab. 5.6 that the real parts of \underline{E}_x and \underline{E}_y for skins match standard values for the elasticity modulus of steel ($\simeq 210$ GPa [Cremer et al. 2005]). Since panel CC is a metallic sandwich panel, its loss factors are much lower than those of panels HC₁ and HC₂. The residual differences on eigenfrequencies and dampings are shown in Fig. 5.17. It can be seen that the uncertainty on the estimation of damping is large $\left\langle \left| \frac{\Delta \alpha_n}{\alpha_n} \right| \right\rangle = 34\%$ and increases with frequency. Since the estimated structural loss factors are very low, the measured modal dampings are very sensitive to the way the panel is suspended

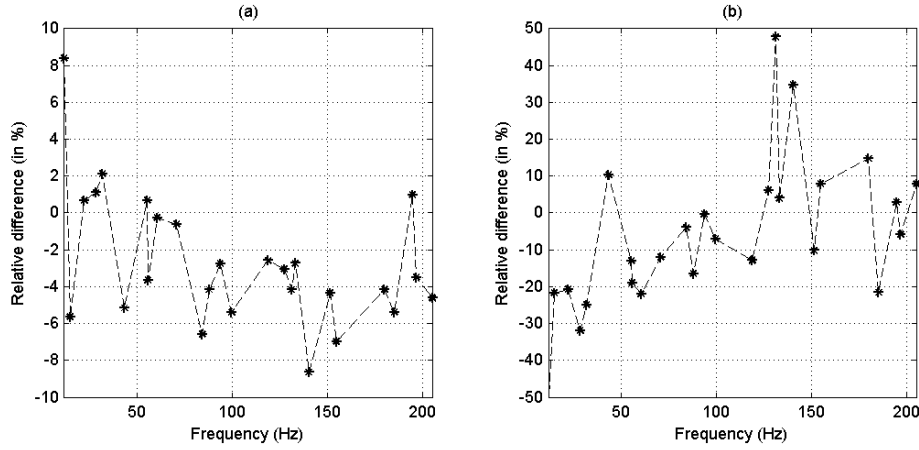


Figure 5.15: Residual differences on eigenfrequencies (a) and dampings (b) for panel HC₂. Numerical values have been obtained with elasticity and damping parameters given in Tab. 5.5.

		E_x	E_y	ν_{xy}	G_{xy}	G_{xz}	G_{yz}
Equivalent plate	Real part	117 GPa	102 GPa	0.33	43 GPa	77 GPa	163 GPa
	Loss factor (%)	0.1	0.1	0.0	0.1	0.0	0.7
Skins	Real part	229 GPa	200 GPa	0.33	84 GPa	-	-
	Loss factor (%)	0.1	0.1	0.0	0.1	-	-
Core	Real part	-	-	-	-	77 GPa	163 GPa
	Loss factor (%)	-	-	-	-	0.0	0.7

Table 5.6: Estimated mechanical parameters for panel CC. Parameters relative the skins and the core are obtained after inversion of Eqs. (5.9) and (5.10)

(thin wires) and to damping due to acoustical radiation. Moreover, these damping mechanisms are also mode-dependant. In the high-frequency range, a systematic discrepancy appears between the measured and the (numerically) modelled damping factors. For the equivalent homogeneous plate, the coincidence frequency f_c is estimated to be approximately 4 kHz. Damping due to acoustical radiation increases as the modal frequency comes close to f_c . In the same spirit, the difference between f_n^{XP} and f_n^{Num} seems to be systematically negative by $\simeq 2\%$. By order of magnitude, this is consistent with air loading in the low frequency range. The relative differences between experimental and numerical modal frequencies remain small $\left\langle \left| \frac{\Delta f_n}{f_n} \right| \right\rangle = 2.3\%$.

Based on the estimated values given in Tab. 5.6, the modal shapes of the extracted modes can be computed by means of the extended Rayleigh-Ritz procedure of Sec. 5.3.3. In the y -direction, panel CC has a maximum of 7 nodal lines in the studied frequency range: its side-length is $4\lambda_y^{\min}$ and $\lambda_y^{\min} = 5.7$ cm, containing $\frac{\lambda_y^{\min}}{s_{\text{cell}}} \simeq 55$ cells. In the x -direction, panel CC has a maximum of 6 nodal lines: $\lambda_x^{\min} = 5$ cm, containing $\frac{\lambda_x^{\min}}{s_{\text{cell}}} \simeq 50$ cells. This is sufficient for the core to be considered as homogeneous. Also, $\frac{2h}{\lambda_{\min}} = 6.6 \times 10^{-2} < 0.08$ in the y -direction, and $\frac{2h}{\lambda_{\min}} = 7.52 \times 10^{-2} < 0.08$ in the y -direction, which is also sufficient for modal frequencies to be predicted by the thick-plate theory. In the frequency

5.3. Paramètres d'élasticité et d'amortissement pertinents pour les plaques "sandwich" épaisses

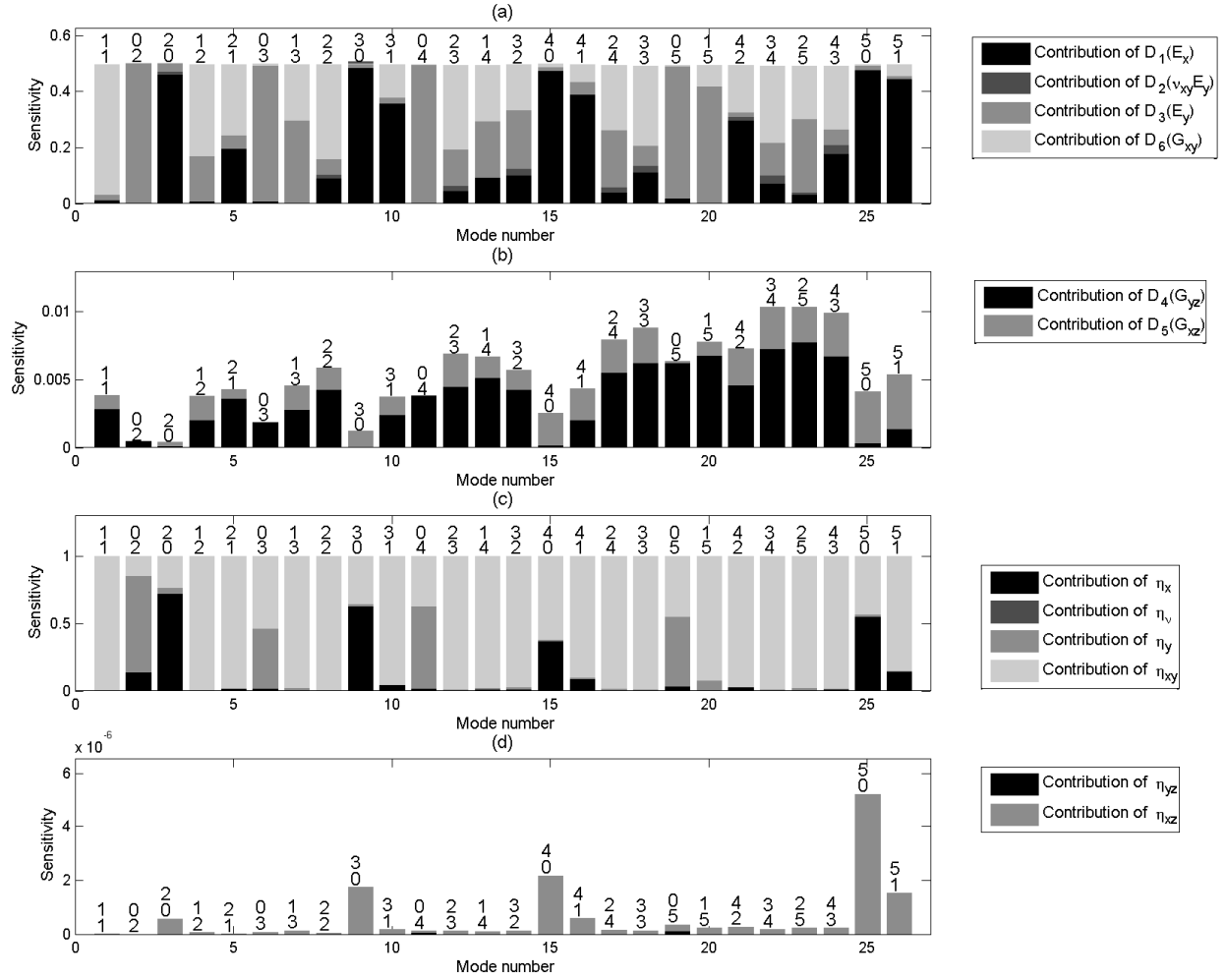


Figure 5.16: Sensitivities of the modal frequencies to the in-plane (a) and out-of-plane (b) elasticity parameters. Sensitivities of the modal dampings to the in-plane (c) and out-of-plane (d) loss factors for panel HC₂. Modes are identified by the numbers of their nodal lines in the x- and y-direction (top and bottom numbers respectively, on top of each bar).

range under consideration, the core of panel CC can be considered as homogeneous and the high modal frequencies are expected to be well predicted by the thick-plate theory.

Results of the sensitivity analysis are presented in Fig. 5.18 for panel CC. It can be seen that sensitivities to the out-of-plane properties are very low compared to sensitivities relative to the in-plane properties and also as compared to sensitivities to the out-of-plane properties of panels HC₁ and HC₂. This means that the modal frequencies and modal dampings are hardly influenced by the out-of-plane complex moduli. As a consequence, the estimations of these parameters must be interpreted very carefully. In this case the core material is too stiff (the last hypothesis listed in Sec. 5.3.2 is not valid): the out-of-plane shear moduli are too high to allow for their precise identification.

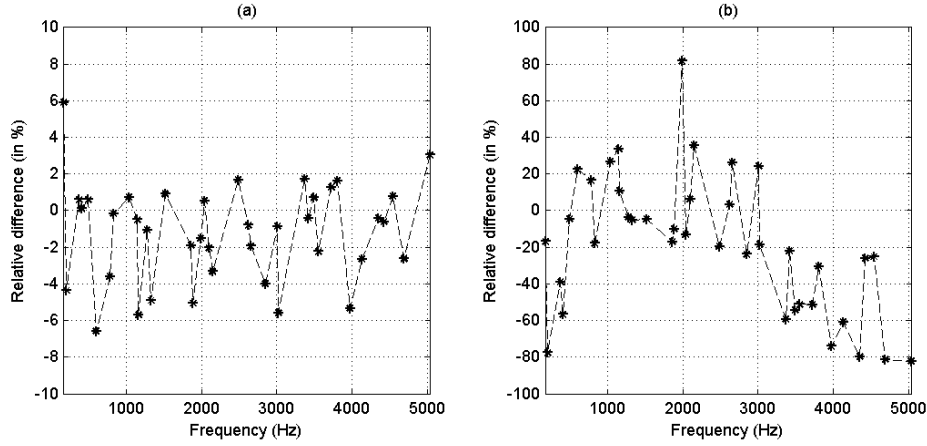


Figure 5.17: Residual difference on eigenfrequencies (a) and dampings (b) for panel CC. Numerical values have been obtained using elasticity and damping parameters given in Tab. 5.6.

5.3.8 Conclusion

In this paper, a method for the measurement of six elasticity moduli and six loss-factors of the constituents of a three-layer symmetrical sandwich material, namely \underline{E}_x^s , \underline{E}_y^s , ν_{xy}^s , \underline{G}_{xy}^s , \underline{G}_{xz}^c , \underline{G}_{yz}^c , has been presented (sandwich structure in the xy -plane). The method directly extends the work of de Visscher et al. [1997] by proposing a means to measure also *out-of-plane* complex moduli. It continues the work of Bastos et al. [2002] with the inclusion of loss factors in the mixed experimental / numerical procedure. Compared to the work of Pagnacco et al. [2007] and to that of Matter et al. [2009, 2011], the present method does not require full-field measurements and is thus much simpler to implement and faster to execute. Compared to the method of Araujo et al. [2009b,a], Barkanov et al. [2009], High Resolution Modal Analysis allows for more modal data to be extracted and used for the estimation of the elastic and damping properties of sandwich materials. Moreover, residuals obtained in the present paper for modal frequencies ($\simeq 1\%$) and dampings ($\simeq 10\%$) are fully in agreement with residuals obtained in [Barkanov et al. 2009, Araujo et al. 2010, Matter et al. 2011] which estimate mechanical parameters of sandwich panels using much less modal data than does the present study.

The method is suited to sandwich structures with heterogeneous cores (*e.g.* honeycomb cores) provided that the panels on which the tests are performed satisfy several geometrical requirements. It has been validated successfully on virtual plates. Results obtained on three sandwich panels suggest that the dynamic behaviour of the sandwich material can be accurately modelled using (1) an equivalent homogeneous plate modelled with first order shear deformation theory, (2) a simple hysteretic model of the type “ $E(1 + j\eta)$ ” for each constituent material and (3) the “light damping” hypothesis for the panel. The consistency of the results with those obtained by static measurements, or on two different panels having the same sandwich structure, also contributes to the validation of the method. The extensive sensitivity analysis combined with the examination of the residual differences left by the optimisation process yields the degree of confidence that can be attributed to the value of each extracted mechanical parameter.

Since this method is simple and needs no heavy experimental apparatus, it is suited to the in-line control of the production of sandwich-materials. This method can also replace classical tensile tests (with some

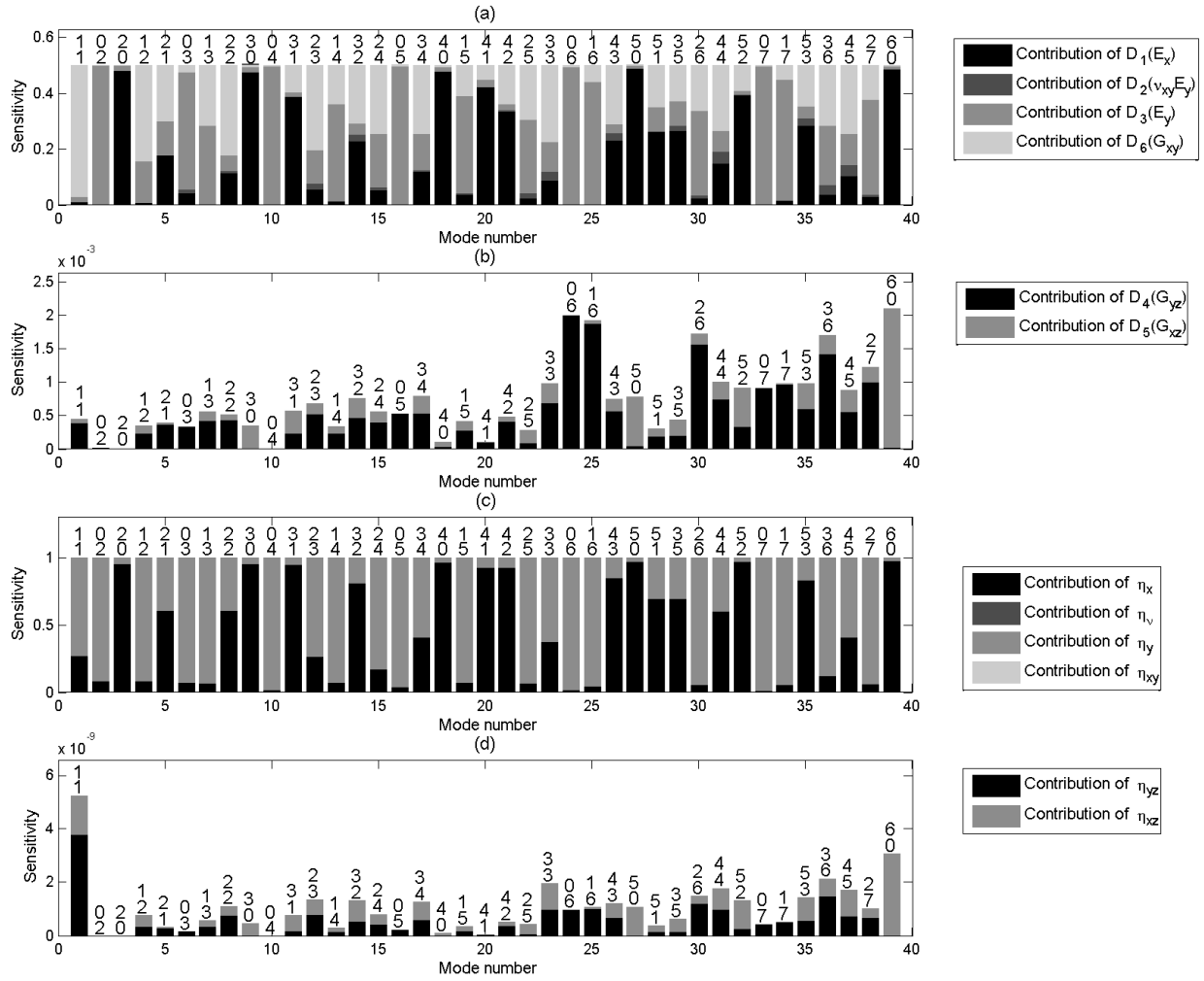


Figure 5.18: Sensitivities of the modal frequencies to the in-plane (a) and out-of-plane (b) elasticity parameters. Sensitivities of the modal dampings to the in-plane (c) and out-of-plane (d) loss factors for panel CC. Modes are identified by the numbers of their nodal lines in the x- and y-directions (top and bottom numbers respectively, on top of each bar).

profit) and also, within the frame of the proposed model, the dynamical mechanical analysis (DMA) for the measurement of elastic and damping material properties.

Acknowledgements

Authors would like to thank Étienne Corteel, Camille Besse, and Dirk Mohr for providing the different test samples that have been characterised in this paper. The authors also thank Kerem Ege for his support regarding High Resolution Modal Analysis and Brian F.G. Katz for English corrections.

5.4 Discussion des hypothèses sur le rayonnement acoustique adoptées pour l’étude dynamique des panneaux

Dans cette section les hypothèses adoptées sur le rayonnement acoustique lors de l’étude réalisée à la section 5.3.7 sont discutées pour chacun des panneaux étudiés expérimentalement : hypothèse de *fluide léger* négligeant l’influence de l’air sur la dynamique du panneau, hypothèse de *faible amortissement* négligeant l’influence de l’amortissement sur les déformées et fréquences modales, et hypothèse selon laquelle les pertes par rayonnement sont négligeables devant les pertes dans le matériau.

5.4.1 Hypothèse de *fluide léger*

Dans l’article présenté à la section 5.3, l’hypothèse dite de *fluide léger* est admise [Junger & Feit 1986, Lesueur 1988, Crighton 1989, Mattei 2007]. Cette hypothèse consiste à négliger les termes faisant intervenir la pression acoustique dans le bilan énergétique régissant la dynamique des panneaux et présenté à la section 5.3.2. Dans le cadre de cette hypothèse, le problème vibratoire et le problème de rayonnement sont ainsi complètement découplés et peuvent donc être étudiés séparément l’un de l’autre. Le but de cette section est de discuter du domaine de validité de cette hypothèse pour les panneaux qui ont été étudiés expérimentalement à la section 5.3.7.

Problème simplifié étudié

Le problème considéré ici est une version simplifiée de la situation expérimentale réelle. Un panneau infini d’épaisseur h fait d’un matériau isotrope de masse volumique ρ_s , de module de Young E et de coefficient de Poisson ν est supposé plongé dans un fluide de masse volumique ρ_a et au sein duquel les ondes de compression se propagent à la vitesse c_a . Les pressions s’exerçant sur les faces supérieure et inférieure du panneau sont notées $p_+(x, y, t)$ et $p_-(x, y, t)$. La différence de pression à laquelle est soumise le panneau est notée $\delta p(x, y, t) = p_+(x, y, t) - p_-(x, y, t)$. Le déplacement transverse du panneau est noté $w(x, y, t)$. Les notations sont résumées à la figure 5.19

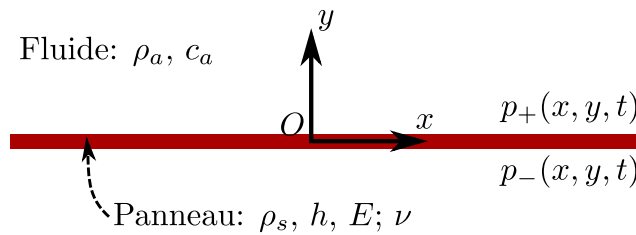


FIG. 5.19 – Notations utilisées pour étudier le domaine de validité de l’hypothèse de *fluide léger*.

La dynamique du panneau, supposé ici par souci de simplicité se comporter comme une plaque *mince*⁴, est régie par l’équation suivante :

$$\rho_s h \frac{\partial^2 w}{\partial t^2} + D \nabla^4 w = \delta p|_{y=0} \quad \text{avec} \quad D = \frac{E h^3}{12(1 - \nu^2)} \quad (5.31)$$

4. Le panneau est supposé satisfaire les hypothèses de Love-Kirchhoff [Cremer et al. 2005]

À l'interface *fluide/solide*, la condition de continuité des vitesses impose de plus :

$$\rho_a \frac{\partial^2 w}{\partial t^2} = \frac{\partial p_+}{\partial y} \Big|_{y=0^+} \quad \text{et} \quad \rho_a \frac{\partial^2 w}{\partial t^2} = \frac{\partial p_-}{\partial y} \Big|_{y=0^-} \quad (5.32)$$

Enfin, la propagation des ondes acoustiques dans le fluide s'écrit :

$$\forall y > 0 : \quad \frac{\partial^2 p_+}{\partial t^2} + c_a^2 \Delta p_+ = 0 \quad \text{et} \quad \forall y < 0 : \quad \frac{\partial^2 p_-}{\partial t^2} + c_a^2 \Delta p_- = 0 \quad (5.33)$$

Recherche de solutions sous forme d'ondes planes

Les solutions de ce problème vibro-acoustique couplé sont ensuite recherchées parmi les ondes planes progressives se propageant dans le panneau selon x :

$$w(x, t) = w \exp [j(k_s x - \omega t)] \quad (5.34a)$$

$$p_+(x, y, t) = p^+ \exp [j(k_x^+ x + k_y y - \omega t)] \quad (5.34b)$$

$$p_-(x, y, t) = p^- \exp [j(k_x^- x - k_y y - \omega t)] \quad (5.34c)$$

La relation de continuité des vitesses sur le plan $y = 0$, Eq. (5.32), implique :

$$k_x = k_s = k_x^+ = k_x^- \quad (5.35)$$

De plus, d'après la propagation des ondes dans le fluide, Eq. (5.33) :

$$k_a^2 = \frac{\omega^2}{c_a^2} = k_x^2 + k_y^2 \quad \text{donc} \quad k_y = \sqrt{\frac{\omega^2}{c_a^2} - k_x^2} = \sqrt{k_a^2 - k_x^2} \quad (5.36)$$

L'injection des solutions sous la forme recherchée dans les équations (5.31) et (5.32) conduit alors à :

$$\frac{k_x^4 - k_v^4}{k_v^4} = -j \frac{2\rho_a}{\rho_s h} \frac{1}{\sqrt{k_a^2 - k_x^2}} \quad \text{avec} \quad k_v^4 = \frac{\rho_s h \omega^2}{D} \quad (5.37)$$

L'équation (5.37) est l'équation de dispersion définissant les nombres d'ondes k_x pouvant se propager dans le panneau et dans le fluide lorsque le problème vibratoire et le problème de rayonnement sont couplés. Le nombre d'onde k_v désigne le nombre d'onde associé aux ondes de flexion se propageant dans le panneau lorsqu'il est plongé dans le vide. Le nombre d'onde k_a désigne le nombre d'onde associé aux ondes de pression se propageant dans le fluide en l'absence de panneau.

Il est important de noter que la pulsation ω intervient par l'intermédiaire de k_a dans la relation (5.37). La pulsation pour laquelle $k_a = k_v$ est appelée *pulsation de coïncidence* et est donnée par :

$$\omega_c = c_a^2 \sqrt{\frac{\rho_s h}{D}} \quad (5.38)$$

En basses fréquences, *i.e.* pour $\omega < \omega_c$, les ondes de flexion se propagent dans le panneau en l'absence de fluide moins vite que les ondes sonores et $k_a < k_v$: on parle alors de *régime subsonique*. L'étude décrite dans la suite est valable uniquement dans ce régime.

Correction apportée aux nombres d’onde de flexion par le couplage avec le fluide

Dans le vide, les ondes de flexion se propagent dans le panneau avec les nombres d’onde :

$$k_x^\alpha = \pm k_v \quad (5.39a)$$

$$k_x^\beta = \pm j k_v \quad (5.39b)$$

Les ondes de flexion dans le panneau en présence de fluide se propagent donc quasiment comme si le panneau était plongé dans le vide si [Mattei 2007] :

$$k_x^\alpha = \pm k_v(1 + \gamma_\alpha) \quad \text{avec } |\gamma_\alpha| \ll 1 \quad (5.40a)$$

$$k_x^\beta = \pm j k_v(1 + \gamma_\beta) \quad \text{avec } |\gamma_\beta| \ll 1 \quad (5.40b)$$

Après report dans l’équation (5.37) et développement limité à l’ordre 1, les ondes de flexion dans le panneau en présence de fluide se propagent donc quasiment comme si le panneau était plongé dans le vide seulement si (*hypothèse de fluide léger*) :

$$\left| \frac{1}{2h\rho_s} \times \frac{\rho_a}{\sqrt{k_a^2 - (k_x^\alpha)^2}} \right| \ll 1 \quad (5.41a)$$

$$\left| \frac{1}{2h\rho_s} \times \frac{\rho_a}{\sqrt{k_a^2 - (k_x^\beta)^2}} \right| \ll 1 \quad (5.41b)$$

Dans la relation (5.41a), le premier terme ($h\rho_s$) correspond à la masse surfacique du panneau étudié. Dans le second terme et dans le cas *subsonique*, l’expression :

$$\frac{1}{\sqrt{k_a^2 - (k_x^\alpha)^2}} = \frac{1}{k_y} = \frac{j\lambda_y}{2\pi} \quad (5.42)$$

fait apparaître λ_y , longueur caractéristique d’atténuation du champ acoustique dans la direction y (orthogonale au panneau). Ainsi, le second terme de l’Eq. (5.41a) peut s’interpréter comme la masse surfacique du “panneau d’air” que déplace avec lui le panneau physique lorsqu’il vibre.

Physiquement, cette relation indique qu’un fluide pourra être considéré comme *léger* par rapport à un panneau donné d’autant mieux que $\frac{\rho_a \lambda_y}{\rho_s h}$ sera petit devant un. Ainsi un panneau *lourd* (fait d’acier par exemple) mis en vibration dans un fluide *léger* (l’air) sera moins influencé par ce dernier que s’il était plongé dans un fluide plus *lourd* (de l’eau par exemple). De plus, plus un panneau est épais par rapport à la longueur d’onde λ_y associée à la propagation de la pression dans la direction y , moins le fluide influence sa dynamique.

Enfin, les relations Eqs. (5.40) sont injectées une à une dans l’Eq. (5.37). En régime subsonique et pour des pulsations suffisamment⁵ loin de ω_c , on trouve alors [Mattei 2007] :

5. Plus précisément, les conditions à vérifier sont : $\frac{k_a^2 - k_v^2}{k_v^2} \gg 2\gamma_\alpha$ et $\frac{k_a^2 + k_v^2}{k_v^2} \gg 2\gamma_\beta$

$$\gamma_1 = \pm \frac{\rho_a}{2\rho_s h} \frac{1}{\sqrt{k_v^2 - k_a^2}} \quad (5.43a)$$

$$\gamma_2 = \pm \frac{j\rho_a}{2\rho_s h} \frac{1}{\sqrt{k_v^2 + k_a^2}} \quad (5.43b)$$

Application aux panneaux étudiés expérimentalement à la section 5.3.7

La validité de l'hypothèse de *fluide léger* est ici étudiée pour les panneaux HC₁, HC₂ et CC présentés à la section 5.3.7 et vibrant dans l'air. Pour chacun de ces panneaux, il est possible de déterminer les écarts γ_α et γ_β induits par l'air sur les nombres d'onde se propageant dans le panneau à l'aide des équations (5.43).

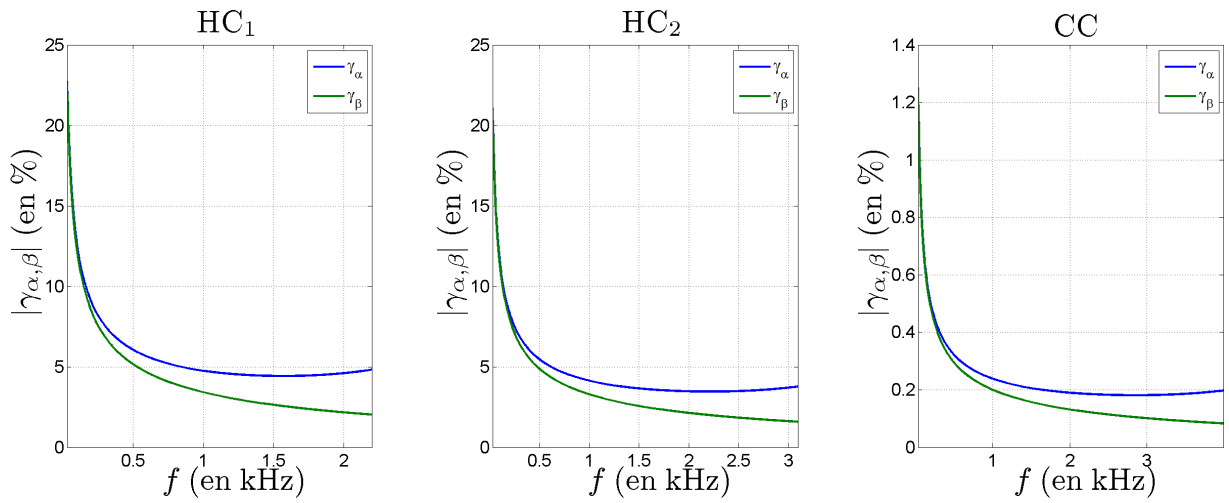


FIG. 5.20 – Valeurs absolues des écarts relatifs γ_α et γ_β (exprimée en %) induits par le couplage des différents panneaux avec l'air.

Pour le panneau HC₁, les valeurs des paramètres choisis sont $h = 5.28$ mm, $\rho_s = 88.97$ kg/m³, $E = 1.22$ GPa, $\nu = 0.3$ et conduisent à $f_c = 2980$ Hz. Pour le panneau HC₂, les valeurs des paramètres choisis sont $h = 4.2$ mm, $\rho_s = 101.6$ kg/m³, $E = 1.11$ GPa, $\nu = 0.3$ et conduisent à $f_c = 4200$ Hz. Pour le panneau CC, les valeurs des paramètres choisis sont $h = 1.88$ mm, $\rho_s = 3363$ kg/m³, $E = 109$ GPa, $\nu = 0.31$ et conduisent à $f_c = 5420$ Hz.

La figure 5.20 présente les valeurs de γ_α et de γ_β obtenues en fonction de la fréquence pour chacun des trois panneaux étudiés expérimentalement à la section 5.3.7. Sur les trois figures, un comportement typique en basse fréquences est observé : pour $f \rightarrow 0$, les nombres d'onde $k_x^{\alpha,\beta}$ s'écartent sensiblement de k_v à cause du phénomène dit de *masse ajoutée*. En effet, plus la fréquence est basse, plus la longueur caractéristique λ_y d'atténuation du champ de pression acoustique va être grande (voir Eq. (5.42)), et donc plus le panneau physique va devoir déplacer avec lui un “panneau d'air” d'épaisseur importante. Cependant, étant donné que des panneaux de dimensions finies et non-bafflés ont été utilisés en pratique dans la section 5.3.7, il est très probable que les effets de “court-circuit acoustique” limitent les effets du couplage en basses fréquences. Le fait que sur les figures 5.12, 5.15, 5.17 aucun écart systématique ne soit observé entre les fréquences propres obtenues expérimentalement et celle calculées numériquement sans prendre en compte le couplage avec l'air argumente en ce sens.

En termes d'ordre de grandeur, les résultats quantitatifs obtenus pour les panneaux HC₁ et HC₂ sont assez

comparables. Pour ces panneaux l’ordre de grandeur de l’imprécision réalisée sur les nombres d’ondes en adoptant l’hypothèse de *fluide léger* est d’environ 3 – 4 %. Du fait de leur relative légèreté par rapport à l’air ces panneaux sont donc quand même influencés par la présence de l’air. Il apparaît donc comme justifié de négliger l’influence de l’air dans les domaines de fréquence étudiés pour les panneaux HC₁ et HC₂ dans l’étude présentée à la section 5.3. En ce qui concerne le panneau CC, qui est bien plus dense que l’air, les imprécisions sont de l’ordre de 0.2 % et donc l’hypothèse de fluide léger est dans ce cas parfaitement justifiée.

5.4.2 Hypothèse de faible amortissement

Dans l’article présenté à la section 5.3, l’hypothèse dite de *faible amortissement* [Geradin & Rixen 1997] est utilisée pour obtenir des approximations des fréquences f_k^{NC} et des déformées ξ_k^{NC} des modes du système non-conservatif \mathcal{P}_{NC} étudié à partir de celles du système conservatif \mathcal{P}_{C} associé. Le but de cette discussion est d’étudier analytiquement ou numériquement la qualité des approximations ainsi réalisées.

Système étudié

Soit \mathcal{P}_{NC} un système non-conservatif possédant N degrés de liberté et dissipant de l’énergie via un amortissement de type *hystérétique* au travers d’une matrice de raideur complexe $\mathbf{K} = \mathbf{K}' + i\mathbf{K}''$ [Pritz 1998, Muravskii 2004]. Si l’on représente l’ensemble des degrés de liberté de ce système au sein du vecteur \mathbf{q} , sa dynamique vibratoire en régime libre est décrite par :

$$(\mathbf{K}' + i\mathbf{K}'')\mathbf{q} + \mathbf{M}\ddot{\mathbf{q}} = 0 \quad (5.44)$$

Les N éléments de la base modale décrivant l’ensemble des solutions de l’équation (5.44) sont alors de la forme $\mathbf{q}_k = \xi_k e^{\lambda_k t}$ et satisfont à l’équation aux valeurs propres suivante :

$$[(\mathbf{K}' + i\mathbf{K}'') + \lambda_k^2 \mathbf{M}] \xi_k = 0 \quad (5.45)$$

Le vecteur propre ξ_k est appelé la *déformée modale* et peut être complexe dans le cas général. Si elle est exprimée dans la base modale $\{\xi_k\}_{k \in [1..N]}$, l’équation (5.44) possédant N degrés de liberté et régissant la dynamique du système étudié peut être scindée en N équations découplées et possédant chacune 1 degré de liberté [Ma et al. 2009].

Cas conservatif

Dans le cas *conservatif*, l’amortissement est nul, *i.e.* $\mathbf{K}'' = 0$. Les valeurs propres λ_k^{C} associées à chacune des déformées modales sont alors imaginaires pures : $\lambda_k^{\text{C}} = i\omega_k^{\text{C}} = i2\pi f_k^{\text{C}}$. De plus, les déformées modales ξ_k^{C} de chaque élément sont réelles [Geradin & Rixen 1997].

Amortissement hystérétique proportionnel

Dans le cas où l’amortissement est supposé *hystérétique proportionnel*, la matrice de raideur complexe prend la forme particulière suivante : $\mathbf{K} = \mathbf{K}'(1 + i\eta)$. Dans ce cas, l’amortissement hystérétique et

la raideur sont répartis dans le système de façon rigoureusement similaire. La matrice d'amortissement $\mathbf{K}'' = \eta \mathbf{K}'$ est ainsi diagonale dans la base modale de \mathcal{P}_C . Les déformées modales ξ_k^C de \mathcal{P}_C forment alors aussi une base orthogonale de \mathcal{P}_{NC} , ce qui conduit à [Sestieri & Ibrahim 1994, Phani & Adhikari 2008] :

$$\xi_k^{NC} = \xi_k^C \quad (5.46)$$

De plus, du fait de l'ajout du terme en $i\eta \mathbf{K}$ dans l'équation de la dynamique, les valeurs propres de \mathcal{P}_{NC} diffèrent de celles de \mathcal{P}_C et s'expriment :

$$\lambda_k^{NC} = i\omega_k^{NC} - \alpha_k = i\omega_k^C \sqrt{1 - \frac{\eta^2}{4}} - \frac{\eta\omega_k^C}{2} \quad (5.47)$$

Il est important de noter qu'une partie de la correction apportée par l'amortissement hystérétique proportionnel à la valeur propre de \mathcal{P}_C prend la forme d'une partie réelle négative et donc conduit bien à des oscillations amorties. L'autre partie, du second ordre en l'amortissement, a pour effet de réduire d'un facteur $\sqrt{1 - \frac{\eta^2}{4}}$ la pulsation des oscillations. Enfin, on constate que ces deux corrections sont proportionnelles à ω_k^C .

Amortissement hystérétique

Plus généralement, si l'amortissement est supposé *hystérétique*, alors la partie complexe $i\mathbf{K}''$ de la matrice de raideur n'est plus nécessairement diagonale quand elle est exprimée dans la base des déformées modales ξ_k^C de \mathcal{P}_C . En utilisant cette représentation, les équations régissant la dynamique des différents modes de \mathcal{P}_{NC} ne sont alors plus découplées les unes des autres. Il est toutefois possible de chercher à évaluer, au premier ordre en l'amortissement, de quelle façon les déformées modales ξ_k^{NC} et les valeurs propres λ_k^{NC} qui découplent effectivement les équations de la dynamique du système \mathcal{P}_{NC} s'éloignent de celles du système \mathcal{P}_C . Les notations suivantes sont alors adoptées [Geradin & Rixen 1997] :

$$\lambda_k^{NC} \simeq i\omega_k^C + \Delta\lambda_k \quad (5.48a)$$

$$\xi_k^{NC} \simeq \xi_k^C + \Delta\xi_k \quad \text{avec} \quad \Delta\xi_k = \sum_{\substack{s=1 \\ s \neq k}}^N \zeta_s \xi_s \quad (5.48b)$$

En réinjectant les formes définies par l'équation (5.48) dans l'équation (5.45) et en négligeant les termes d'ordre 2, on obtient :

$$[i\mathbf{K}'' + 2i\omega_k^C \Delta\lambda_k \mathbf{M}] \xi_k + (\mathbf{K}' - (\omega_k^C)^2 \mathbf{M}) \Delta\xi_k \simeq 0 \quad (5.49)$$

La multiplication de l'équation (5.49) à gauche par la transposée de ξ_k^C conduit alors à l'expression suivante de $\Delta\lambda_k$:

$$\Delta\lambda_k \simeq -\frac{\beta_{kk}\omega_k^C}{2} \quad \text{avec} \quad \beta_{kk} = (\xi_k^C)^T \mathbf{K}'' \xi_k^C \quad (5.50)$$

La correction au premier ordre apportée par l’amortissement à la valeur propre λ_k^{NC} prend donc la forme d’une partie réelle négative et conduit bien encore une fois à des oscillations amorties. Cette correction fait intervenir uniquement les termes diagonaux de la matrice d’amortissement \mathbf{K}'' et a une forme très similaire à la correction apportée par un amortissement hystérétique proportionnel présentée dans l’équation (5.47). Les termes d’ordre 2 étant négligés, le terme correctif trouvé pour l’amortissement hystérétique proportionnel et conduisant à une diminution de la pulsation des oscillations dans l’équation (5.47) n’est pas visible ici.

La multiplication de l’équation (5.49) à gauche par les transposées des éléments de la base modale associée au système non-amorti $\{\xi_l^{\text{C}}\}_{l \in [1..N], l \neq k}$ conduit de plus à l’expression suivante de $\Delta \xi_k$:

$$\Delta \xi_k \simeq \sum_{\substack{s=1 \\ s \neq k}}^N \frac{i\beta_{sk}}{\mu_s ((\omega_k^{\text{C}})^2 - (\omega_s^{\text{C}})^2)} \xi_s^{\text{C}} \quad \text{avec} \quad \beta_{sk} = (\xi_l^{\text{C}})^{\text{T}} \mathbf{K}'' \xi_k^{\text{C}} \quad \text{et} \quad \mu_k = (\xi_k^{\text{C}})^{\text{T}} \mathbf{M} \xi_k^{\text{C}} \quad (5.51)$$

Cette approximation au premier ordre de $\Delta \xi_k$ fait apparaître uniquement les termes non-diagonaux β_{sk} de la matrice d’amortissement \mathbf{K}'' exprimée dans la base modale de \mathcal{P}_{C} . La correction apportée par l’amortissement est imaginaire pure, et de l’ordre de grandeur de β_{sk} . De plus, si \mathbf{K}'' est diagonale dans la base modale de \mathcal{P}_{C} (amortissement *hystérétique proportionnel*), alors les coefficients β_{sk} sont nuls et l’on retrouve bien $\Delta \xi_k = 0$, ce qui est cohérent avec l’équation (5.46).

Illustration numérique

En pratique, l’amortissement peut être *hystérétique*, mais est rarement *proportionnel*. Les corrections apportées par ce type d’amortissement n’étant pas calculables analytiquement, elles ont donc été étudiées numériquement sur une plaque orthotrope homogène virtuelle à titre d’illustration.

Les caractéristiques géométriques et élastiques de la plaque étudiée sont données dans la table 5.7. L’amortissement *hystérétique*, réparti sur l’ensemble de la plaque, est défini par $\eta = \eta_x = \frac{\eta_y}{2} = \frac{\eta_{xy}}{3} = 2\eta_{\nu_{xy}} = \frac{\eta_{xz}}{4} = \frac{\eta_{yz}}{4}$ et le coefficient η prend des valeurs variant entre $\eta = 1 \%$ et $\eta = 20 \%$. Les valeurs des fréquences f_k^{C} et des déformées modales ξ_k^{C} liées au système conservatif et celles f_k^{NC} et ξ_k^{NC} liées au système non-conservatif ont été calculées avec 20 éléments par côtés pour les 23 premiers modes de cette plaque virtuelle en utilisant la procédure numérique décrite à la section 5.3.

l_x (m)	l_y (m)	h (m)	ρ (kg/m ³)
0.4	0.6	5×10^{-3}	700

E_x (GPa)	E_y (GPa)	ν_{xy}	G_{xy} (GPa)	G_{xz} (GPa)	G_{yz} (GPa)
2	5	0.33	1	1	1

TAB. 5.7 – Paramètres géométriques et élastiques de la plaque orthotrope homogène virtuelle étudiée.

Sur la figure 5.21 sont présentés les résultats numériques donnant les parties réelle et imaginaire de $\epsilon_k = \frac{\Delta \lambda_k}{\omega_k^{\text{C}}}$ en fonction de la valeur η prise par le coefficient η déterminant l’amortissement *hystérétique*,

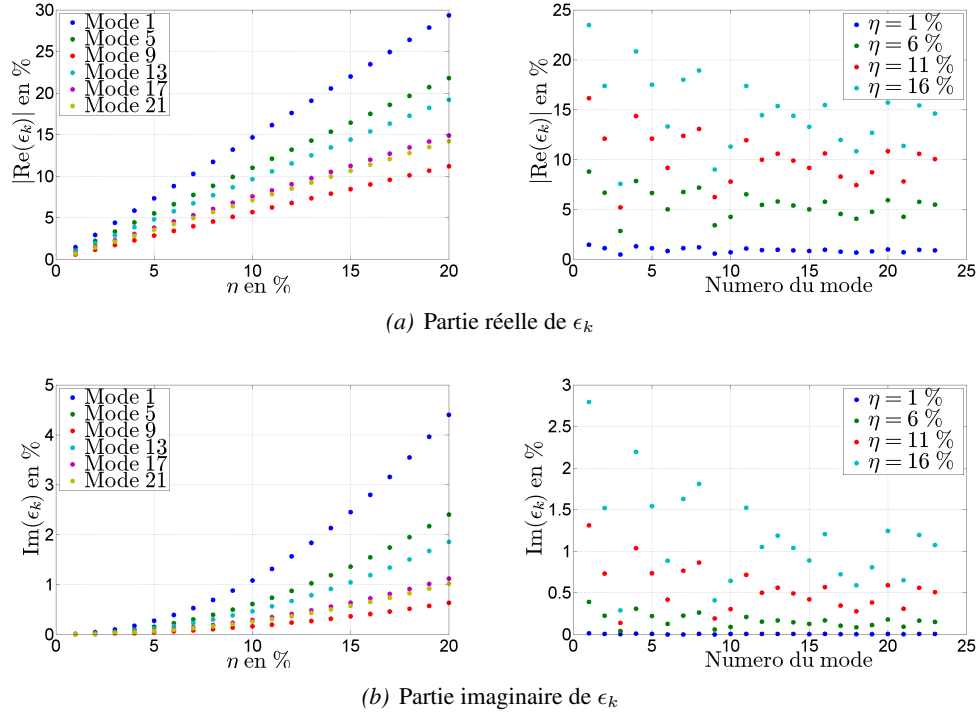


FIG. 5.21 – Résultats numériques donnant les parties réelle et imaginaire de $\epsilon_k = \frac{\Delta\lambda_k}{\omega_k^C}$ en fonction de la valeur du paramètre η définissant l'amortissement *hystérétique*, et du mode considéré.

et du mode considéré. L'étude des courbes donnant $|\text{Re}(\epsilon_k)|$ en fonction de η révèle que dans ce cas où l'amortissement n'est pas proportionnel la correction apportée à la partie réelle est bien d'ordre 1 en amortissement, conformément à l'équation (5.50). L'étude des courbes donnant $\text{Im}(\epsilon_k)$ confirme par ailleurs que la correction apportée à la partie imaginaire est d'ordre 2 en amortissement, comme dans le cas où l'amortissement est proportionnel. Toujours dans le cas de l'amortissement proportionnel, l'approche analytique prédit que les parties réelle et imaginaire de ϵ_k sont indépendantes du mode considéré. L'étude de $|\text{Re}(\epsilon_k)|$ et de $\text{Im}(\epsilon_k)$ en fonction du mode considéré révèle que, pour un amortissement non-proportionnel, ceci est approximativement vrai pour les faibles valeurs d'amortissement (*i.e.* $\eta < 5\%$).

Sur la figure 5.22 est présentée la norme de l'écart relatif $\zeta_k = \frac{\|\Delta\xi_k\|}{\|\xi_k^{\text{NC}}\|}$ entre les déformées modales du système conservatif ξ_k^C et celles du système non-conservatif ξ_k^{NC} . Dans le cas étudié ici, la correction apportée par l'amortissement aux déformées modales du système conservatif est vraiment d'un très petit ordre de grandeur comparée aux corrections apportées aux valeurs propres, même pour des valeurs du paramètre η relativement importantes. De plus, les déformées modales des premiers modes semblent plus affectées par l'amortissement que celles des modes d'ordre élevé.

Conclusion quand à l'hypothèse de faible amortissement

Cette étude analytique et cette illustration numérique montrent que dans le cas d'un amortissement *hystérétique* faible, les fréquences propres f_n^C du système conservatif et celles f_n^{NC} du système non-conservatif ne diffèrent que d'un terme à l'ordre 2 en l'amortissement. D'autre part les déformées modales ξ_k^C du sys-

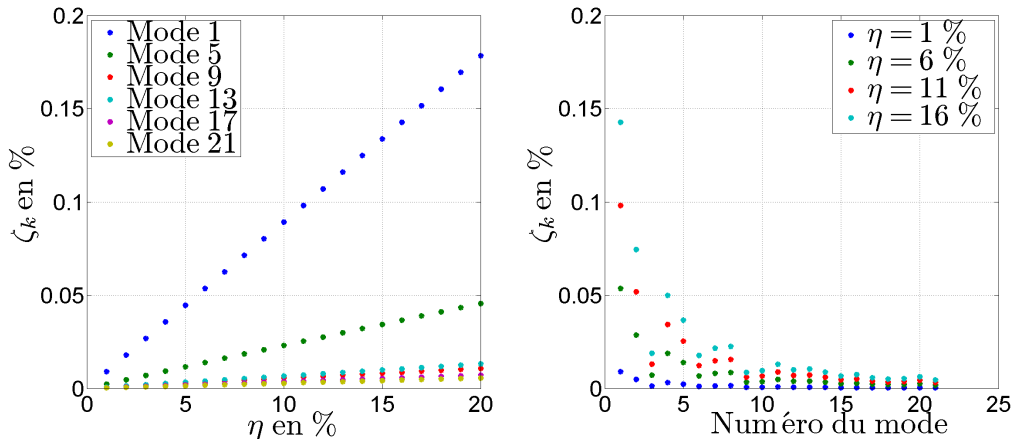


FIG. 5.22 – Norme de l'écart relatif $\zeta_k = \frac{\|\Delta \xi_k\|}{\|\xi_k^C\|}$ entre les déformées modales du système conservatif ξ_k^C et du système non-conservatif ξ_k^{NC} .

tème conservatif et celles ξ_k^{NC} du système non-conservatif ne diffèrent que d'un terme du même ordre de grandeur que les coefficients de couplage inter-modal dus à l'amortissement, qui sont généralement petits.

Lors de l'étude de plaques dissipant de l'énergie par un mécanisme d'amortissement *hystérétique* et ayant des facteurs d'amortissement de l'ordre de 5 % l'hypothèse de *faible amortissement* qui consiste à dire que les fréquences et déformées modales du système conservatif sont les mêmes que celles du systèmes non-conservatif est donc justifiée et permet d'obtenir des approximations valables au premier ordre des fréquences f_k^{NC} et des déformées ξ_k^{NC} des modes du système non-conservatif \mathcal{P}_{NC} .

5.4.3 Majoration des pertes par rayonnement

Dans la section 5.3, il est supposé que le phénomène de dissipation d'énergie majoritaire est l'amortissement structural au sein du matériau composant le panneau. Or les pertes sont en réalité dues à différentes sources d'amortissement : pertes structurales dans le matériau, pertes par rayonnement acoustique à l'infini, pertes dues à la fixation du panneau, etc. L'objectif de cette partie est donc d'obtenir une majoration du facteur d'amortissement associé au rayonnement acoustique au sein des panneaux qui ont été étudiées expérimentalement à la section 5.3.7. Cette majoration sera ensuite comparée aux facteurs d'amortissement effectivement mesurés pour chacun des panneaux. Cela permettra de discuter l'hypothèse selon laquelle la majorité des pertes mesurées expérimentalement sont dues aux phénomènes de dissipation ayant lieu dans le matériau.

Facteur d'amortissement associé au rayonnement acoustique d'un panneau bafflé

Dans un système dissipatif, il est possible d'interpréter le facteur d'amortissement $\alpha(\omega)$ à partir de la proportion de l'énergie totale stockée qui est dissipée au cours d'un cycle [Ungar & Kerwin 1962, Cremer et al. 2005] :

$$\alpha(\omega) = \frac{\omega}{2\pi} \frac{E_{\text{diss}}}{E_{\text{tot}}} \quad (5.52)$$

Considérons le système dissipatif formé d'un panneau de dimensions finies dont le déplacement transverse en chaque point est connu. Si l'hypothèse est faite que ce panneau est *bafflé*, alors la puissance $\Pi(\omega)$ rayonnée par le panneau bafflé en champ lointain est aisément calculable [Wallace 1972, 1987, Berry et al. 1990, Anderson & Bratos-Anderson 2005].

L'énergie $E_{\text{diss}}^{\text{ray}}$ dissipée par rayonnement acoustique à l'infini par le panneau bafflé s'écrit :

$$E_{\text{diss}}^{\text{ray}} = \frac{2\pi}{\omega} \Pi(\omega) \quad (5.53)$$

Le facteur d'amortissement $\alpha^{\text{ray}}(\omega)$ associé au rayonnement acoustique à l'infini du panneau bafflé est donc :

$$\alpha^{\text{ray}}(\omega) = \frac{\Pi(\omega)}{E_{\text{tot}}} \quad (5.54)$$

avec E_{tot} l'énergie cinétique du panneau qui est connue ici.

Application aux panneaux étudiés expérimentalement

L'utilisation de la méthode de Rayleigh-Ritz présentée à la section 5.3.3 permet de déterminer le déplacement transverse en chaque point des panneaux étudiés expérimentalement pour chacun de leurs modes propres. En utilisant la démarche exposée précédemment, il est alors possible d'estimer les facteurs d'amortissement dus au rayonnement $\alpha^{\text{ray}}(\omega_n)$ pour chacun des modes des panneaux.

Les panneaux étudiés expérimentalement étant non-bafflés et les calculs précédents étant effectués sur des panneaux bafflés, les valeurs trouvées pour $\alpha^{\text{ray}}(\omega_n)$ ne sont cependant que des majorations des amortissements modaux réellement liés au rayonnement acoustique à l'infini [Atalla et al. 1996, Laulagnet 1998, Putra & Thompson 2010]. Cette majoration peut cependant permettre de déterminer dans quels domaines fréquentiels les amortissements modaux liés au rayonnement sont effectivement négligeables devant ceux mesurés expérimentalement.

La figure 5.23 présente les amortissements modaux mesurés pour chacune des plaques étudiées expérimentalement à la section 5.3.7 ainsi que la part prise par le rayonnement dans les amortissements modaux pour un panneau bafflé. Sur la figure 5.23(a) sont présentés les résultats correspondant au panneau HC₁. Pour les 14 premiers modes (*i.e.* pour $f < 400$ Hz) la part du rayonnement dans les amortissements modaux (dans cas où le panneau est bafflé) est environ dix fois inférieure aux amortissements modaux effectivement mesurés (sur un panneau non-bafflé). Pour ces 14 premiers modes il est donc justifié de négliger le rôle du rayonnement dans les phénomènes dissipatifs. Au delà de 400 Hz, la part prise par le rayonnement dans l'hypothèse *bafflé* devient du même ordre de grandeur que les amortissements mesurés expérimentalement dans le cas *non-bafflé*. Il est donc impossible d'affirmer ou d'infirmer l'hypothèse pour ces modes là. Sur la figure 5.23(c) sont présentés les résultats correspondant à le panneau CC. Leur interprétation qualitative est similaire à celle effectuée pour panneau HC₁. Sur la figure 5.23(b) sont présentés les résultats correspondant au panneau HC₂. Pour ce panneau la part du rayonnement dans les amortissements modaux (dans cas où le panneau est bafflé) est toujours largement inférieure aux amortissements modaux effectivement mesurés (sur un panneau non-bafflé). Pour le panneau HC₂ il est donc justifié de négliger les pertes dues au rayonnement lors de l'estimation des facteurs d'amortissement.

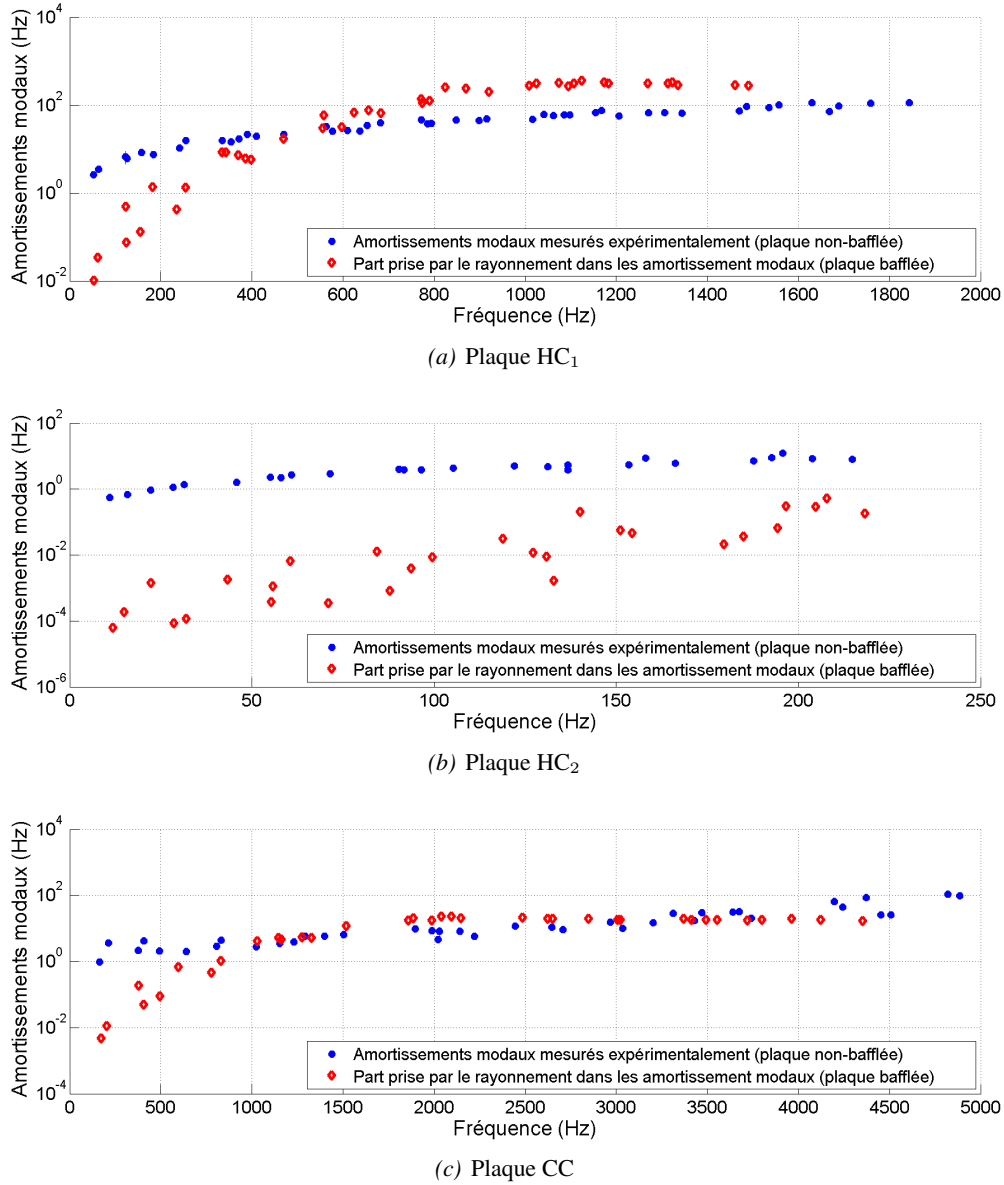


FIG. 5.23 – Comparaison pour les trois plaques étudiées expérimentalement à la section 5.3.7 entre les amortissement modaux *de rayonnement* issus d’un calcul numérique pour un panneau bafflé et des amortissements modaux *expérimentaux* obtenus sur un panneau libre.

Bilan concernant les pertes dues au rayonnement

Dans cette section est discutée l’hypothèse selon laquelle la majorité des pertes mesurées expérimentalement est due aux phénomènes de dissipation ayant lieu dans le matériau composant les panneaux. La méthodologie utilisée permet seulement d’obtenir une majoration de la part prise par le rayonnement à l’infini dans les pertes. Cela ne permet donc pas forcément de conclure de façon ferme. Il est cependant possible de dire que l’hypothèse est complètement justifiée pour le panneau HC₂ et justifiée seulement pour les 10 – 15 premiers modes des panneaux HC₁ et CC. Pour les autres modes de ces deux panneaux, nous ne disposons pas d’arguments permettant de confirmer ou d’infirmier cette hypothèse. Une façon de lever cette indétermination et d’estimer la part relative prise par le rayonnement dans les amortissements

modaux serait d'utiliser les procédures numériques qui ont été développées pour les plaques non-bafflées [Atalla et al. 1996, Nelisse et al. 1998, Laulagnet 1998]. Ces procédures sont cependant plus complexes à mettre en œuvre et plus coûteuses en termes de temps de calcul.

5.5 Conclusion

Ce chapitre présente et valide une méthode permettant l'estimation des paramètres mécaniques nécessaires à la description de la dynamique des panneaux de matériau sandwich composant le SMART-I². Cette méthode ne nécessite qu'une seule mesure dynamique, ponctuelle, réalisée sur une plaque taillée dans le matériau à décrire. Elle permet l'estimation des paramètres d'élasticité et d'amortissement du matériau testé, relatifs à une dynamique de plaque. Ces paramètres sont estimés par le biais d'une procédure mixte numérique/expérimentale consistant à mettre à jour les paramètres d'un modèle numérique en cherchant à minimiser l'écart entre les données obtenues expérimentalement et les données prédites par ce modèle. De plus, les hypothèses de *fluide léger*, de *faible amortissement* et de *pertes par rayonnement négligeables*, faites pour les panneaux étudiés expérimentalement, sont raisonnablement validées.

En guise de perspective à la méthode d'estimation des paramètres mécaniques de panneaux de matériau sandwich décrite à la section 5.3, il est possible de chercher à étudier plus finement l'effet des inhomogénéités introduites par les nids d'abeille sur le comportement dynamique des panneaux étudiés ici. Tant que les longueurs d'ondes étudiées sont très grandes devant les dimensions des inhomogénéités, les vibrations «*ne voient pas les nids d'abeille*». Toutefois les nids d'abeille ont quand même tendance à induire un léger comportement non-déterministe dans la dynamique des panneaux étudiés à mesure que la fréquence augmente [Weaver 1989, Bohigas et al. 1991]. Ceci se traduit par le fait que des modes de fréquences propres proches auront une légère tendance à se *repousser* mutuellement. Les homogénéités ont aussi un effet sur la distribution statistique des amortissements modaux : ces effets n'ont cependant été étudiés pour l'instant que d'un point de vue numérique [Burkhardt & Weaver 1996].

Il est ici assez délicat de faire des statistiques concernant ce point étant donné le peu d'informations modales disponibles. Une perspective intéressante pourrait alors être d'essayer d'obtenir, grâce à l'analyse modale haute-résolution [Ege et al. 2009], plus de données expérimentales (fréquences modales et amortissement modaux) sur des panneaux contenant des inhomogénéités. Ceci permettrait d'évaluer expérimentalement l'influence des inhomogénéités sur les répartitions des écarts inter-modaux et des amortissements modaux et de comparer les résultats expérimentaux, numériques et théoriques.

Chapitre 6

Caractérisation expérimentale des “Large Multi-Actuator Panels” (LaMAPs)

Sommaire

6.1	Introduction	157
6.2	Modélisation simple des LaMAPS	158
6.2.1	Modèle de plaque mince homogène, isotrope et infinie	158
6.2.2	Grandeurs physiques caractéristiques de la dynamique des LaMAPs	158
6.2.3	Action des excitateurs sur le panneau	160
6.3	Confrontation du modèle à l’expérience	161
6.3.1	Protocole expérimental	161
6.3.2	Analyse temporelle	163
6.3.3	Analyse fréquentielle	169
6.4	Conclusion et perspectives	173

6.1 Introduction

L’établissement d’un modèle dynamique complet des deux “*Large Multi-Actuator Panels*” (LaMAPs) composant le SMART-I² suppose l’étude et la modélisation de nombreux phénomènes complexes à décrire individuellement, mais plus encore à coupler les uns aux autres. Il est possible de citer par exemple :

- les conditions aux limites
- la diffraction induite par la présence des excitateurs
- les interactions mécaniques et acoustiques entre les deux panneaux

Ce chapitre présente une *étude préliminaire* ayant pour but d’apporter un éclairage expérimental sur ces différents phénomènes en comparant des résultats expérimentaux obtenus sur les LaMAPs aux prédictions d’un modèle dynamique simple. La section 6.2 présente le modèle dynamique simple utilisé pour décrire le comportement vibratoire des LaMAPs (modèle de plaque mince, homogène, isotrope, infinie et excitation ponctuelle) ainsi que les ordres de grandeurs de différents paramètres caractéristiques de la dynamique du dispositif. La section 6.3 est consacrée aux conditions aux limites, aux interférences dues à la diffraction par les excitateurs voisins, et aux interactions entre les deux LaMAPs.

6.2 Modélisation simple des LaMAPs

Dans cette section, le modèle retenu pour prédire *en première approche* le comportement dynamique des LaMAPs est brièvement présenté accompagné de quelques grandeurs physiques caractéristiques du comportement dynamique de ces panneaux.

6.2.1 Modèle de plaque mince homogène, isotrope et infinie

Le modèle retenu est un modèle de plaque mince, homogène, isotrope et infinie [Cremer et al. 2005]. La propagation d'ondes de flexion dans ce type de plaque, d'épaisseur h , de module d'Young complexe $\underline{E} = E(1 + j\eta)$, de coefficient de Poisson ν , et de masse volumique ρ est déterminée par deux paramètres :

– la masse surfacique :

$$\rho h \quad \text{en kg/m}^2 \quad (6.1)$$

– la rigidité dynamique complexe $\overline{D} = \overline{D}_0(1 + j\eta_0)$:

$$\overline{D} = \frac{\underline{E}h^3}{12(1 - \nu^2)} \frac{1}{\rho h} \quad \text{en m}^4/\text{s}^2 \quad (6.2)$$

L'équation de dispersion liant le nombre d'onde k et la pulsation ω des ondes de flexion se propageant dans ce type de plaque dépend uniquement de \overline{D} :

$$\overline{D}k^4 = \omega^2 \quad (6.3)$$

Les deux panneaux du SMART-I² sont composés du même matériau que la plaque HC₂ (voir section 5.3.7). Leurs paramètres géométriques, massiques et mécaniques sont rappelés dans les tables 6.1 et 6.2.

Comme le montre la table 6.2, les panneaux du SMART-I² sont composés d'un matériau assez peu anisotrope. Il est donc raisonnable, en première approximation, de faire l'hypothèse que les LaMAPs se comportent comme des matériaux isotropes de module de Young complexe moyen $\underline{E} = 0.5(\underline{E}_x + \underline{E}_y)$, soit $E = 1.1$ GPa et $\eta = 1\%$. Les deux paramètres régissant le comportement dynamique du SMART-I² sont ainsi :

$$\overline{D} = \overline{D}_0(1 + j\eta_0) = 17.7(1 + j0.01) \text{ m}^4/\text{s}^2 \quad \text{et} \quad \rho h = 424 \text{ g/m}^2 \quad (6.4)$$

l_x (m)	l_y (m)	h (m)	ρ (kg/m ³)
2.6	2	4.2×10^{-3}	101

TAB. 6.1 – Paramètres géométriques et massiques des LaMAPs.

6.2.2 Grandeurs physiques caractéristiques de la dynamique des LaMAPs

À partir de la paramétrisation effectuée à la section précédente, il est possible de calculer différentes grandeurs physiques permettant de mieux appréhender le comportement dynamique des LaMAPs.

	\underline{E}_x	\underline{E}_y	$\underline{\nu}_{xy}$	\underline{G}_{xy}
Partie réelle	1 GPa	1.2 GPa	0.33	0.48 GPa
Facteur d'amortissement (%)	1	1	0	3.2

TAB. 6.2 – Paramètres mécaniques des LaMAPs.

Fréquence de coïncidence

En premier lieu, on peut s'intéresser à la fréquence de coïncidence f_c , déjà évoquée à la section 5.4.1, qui caractérise la fréquence à partir de laquelle le panneau rayonne efficacement dans l'air. Si la célérité des ondes acoustiques dans l'air est $c_a = 340$ m/s, alors f_c est donnée par :

$$f_c = \frac{c_a^2}{2\pi \overline{D}_0^{1/2}} \quad \text{soit ici : } f_c = 4.3 \text{ kHz} \quad (6.5)$$

Ce type de panneau constitue donc une source acoustique non-directive pour les fréquences inférieures à 4.3 kHz.

Limite fréquentielle jusqu'à laquelle l'approche modale est adaptée

La répartition des modes en fonction de la fréquence au sein des panneaux peut aussi être étudiée. Le premier mode propre du panneau se situe aux alentours de 1.8 Hz et la densité modale asymptotique d'un LaMAP est donnée par :

$$n_\infty = \frac{l_x l_y}{2\overline{D}_0^{1/2}} \quad \text{soit : } n_\infty = 0.6 \text{ modes/Hz} \quad (6.6)$$

La façon dont les différents modes du panneau participent, à une fréquence f donnée, à la réponse du panneau est aussi informative. Le paramètre décrivant ce phénomène est le recouvrement modal μ et peut être approximé par $\mu = \eta_0 f n_\infty$. La fréquence $f_{\mu=1}$ pour laquelle ce recouvrement modal est unitaire est caractéristique de la fréquence à partir de laquelle l'approche modale est inadaptée¹. Cette fréquence est donnée par :

$$f_{\mu=1} = \frac{1}{\eta_0 n_\infty} \quad \text{soit, pour les LaMAPs : } f_{\mu=1} = 145 \text{ Hz} \quad (6.7)$$

Au delà de $f_{\mu=1} = 145$ Hz, le panneau devient donc un système qui ne peut plus être décrit efficacement par une approche modale.

Fréquence où l'atténuation spatiale est comparable à une dimension caractéristique du LaMAP

Lorsqu'une onde de flexion se propage dans une plaque infinie à la fréquence f et selon x , son amplitude s'atténue en $\exp(-\mathcal{I}(k)x) = \exp(-2\pi x/\lambda_{\text{att}})$. $\mathcal{I}(k)$ est la partie imaginaire du nombre k déterminé

1. Cette fréquence est comparable à la fréquence de Schroeder définissant la limite entre comportement modal et comportement diffus en acoustique des salles, en fonction (entre autres) du volume de la salle et de son temps de réverbération [Bruneau & Scelo 2006].

par l'équation de dispersion (6.3). $\lambda_{\text{att}} = 2\pi/\mathcal{J}(k)$ est une longueur typique d'atténuation spatiale due à l'amortissement structurel et dépendant de la fréquence.

Une dernière grandeur caractéristique intéressante est donc la fréquence f_{lc} où la longueur λ_{att} typique de l'atténuation spatiale due à l'amortissement structurel est égale à une dimension caractéristique l_c du panneau étudié, par exemple $l_c = \sqrt{l_x l_y}$. En utilisant l'équation (6.3), on trouve que cette fréquence est donnée par :

$$f_{lc} = 2\pi \overline{D}_0^{-1/2} \frac{4}{\eta_0 \sqrt{l_x l_y}} \quad \text{soit, ici : } f_{lc} \simeq 600 \text{ kHz} \quad (6.8)$$

Pour $f \ll f_{lc} \simeq 600 \text{ kHz}$, l'amortissement structurel n'induit donc pas d'atténuation spatiale importante à l'échelle du LaMAP étudié.

6.2.3 Action des excitateurs sur le panneau

Le panneau étant modélisé par une plaque mince, homogène, isotrope et infinie, il reste maintenant à modéliser l'action des excitateurs sur ce dernier. La zone de contact entre les excitateurs et les panneaux est composée de trois sections réparties sur un anneau de diamètre moyen $d_{\text{moy}} = 3.5 \text{ cm}$ et de largeur 1 cm (voir figure 6.1).



FIG. 6.1 – Faces libre et collée des excitateurs composant les LaMAPs.

La fréquence à laquelle la longueur d'onde est égale à d_{moy} est $f_{\text{exc}} = 2160 \text{ Hz}$. En dessous de cette fréquence, les longueurs d'onde des ondes de flexion engendrées par les excitateurs dans le panneau sont supérieures à l'étendue de la zone de contact des excitateurs avec le panneau. Il est alors raisonnable de supposer en première approximation que les excitateurs ont une action ponctuelle sur les panneaux pour $f \ll f_{\text{exc}} = 2160 \text{ Hz}$.

La réponse en vitesse $V(k, r)$ d'une plaque mince homogène, isotrope, infinie (conditions de Sommerfeld), excitée à la pulsation ω par une excitation ponctuelle, est donnée par [Cremer et al. 2005] :

$$V(k, r) = V_0 \left[H_0^{(2)}(kr) - H_0^{(2)}(-jkr) \right] \quad (6.9)$$

avec

- r la distance au point d'excitation.
- k le nombre d'onde associé aux ondes de flexion se propageant dans la plaque à la fréquence ω et donné par l'équation (6.3).

- $H_0^{(2)}$ la fonction de Hankel de seconde espèce et d'ordre 0.

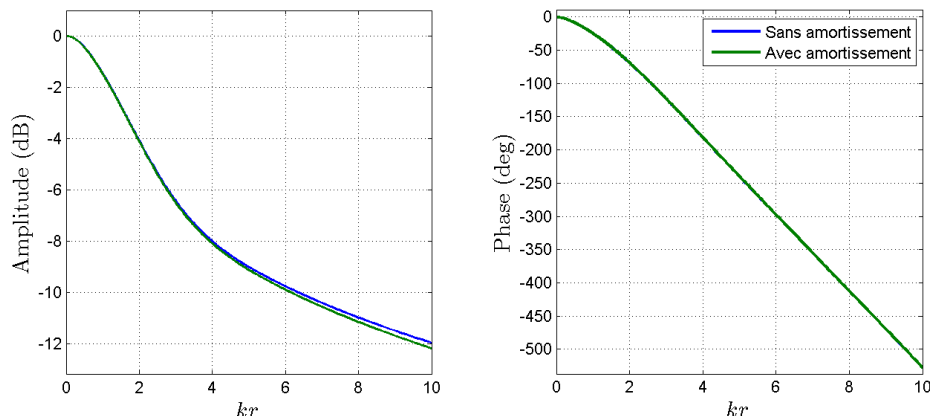


FIG. 6.2 – Amplitude et phase de la réponse en vitesse $V(k, r)$ d'une plaque mince homogène, isotrope, infinie (conditions de Sommerfeld), excitée en $r = 0$ à la pulsation ω par une excitation ponctuelle.

Sur la figure 6.2 sont tracés l'amplitude et la phase de la réponse en vitesse $V(k, r)$ du panneau en prenant en compte ou non l'amortissement structurel. Dans un premier temps, on constate que l'amplitude de la réponse en vitesse décroît asymptotiquement en $1/\sqrt{kr}$ (pente de -6 dB/décade). C'est une conséquence du fait que la structure étudiée est bidimensionnelle et donc que l'énergie transitant au travers de tout cercle centré sur le point d'excitation est constante. D'autre part on constate que, comme il a déjà été mentionné dans le paragraphe précédent, l'amortissement structurel existant dans les panneaux ($\eta_0 = 1\%$) a une très faible influence sur la décroissance spatiale du profil de vitesse. Dans les plaques minces infinies et même en l'absence d'amortissement, l'énergie cinétique est donc naturellement localisée autour du point d'excitation.

6.3 Confrontation du modèle à l'expérience

Dans cette section, les prédictions faites par le modèle simple décrit à la section 6.2 sont confrontées aux résultats expérimentaux obtenus sur les LaMAPs en vue d'obtenir des renseignements concernant les conditions aux limites, les interférences dues à la diffraction par les excitateurs voisins et les interactions entre les LaMAPs.

6.3.1 Protocole expérimental

Toutes les mesures présentées ici ont été réalisées à l'aide d'un vibromètre LASER (Polytec PDV 100) placé à 1.6 m des panneaux. Cet appareil permet de mesurer par effet Doppler la composante parallèle au rayon laser du vibromètre de la vitesse des panneaux en un point donné de leur surface. Le vibromètre a été placé de façon à toujours avoir une incidence normale sur le panneau. Les réponses impulsionnelles correspondant à la vitesse normale des panneaux ont été obtenues en différents points en utilisant la méthode présentée au chapitre 4 avec les paramètres précisés dans la table 6.3.

Les résultats d'une mesure sont présentés à titre d'exemple sur la figure 6.3. Sur cette figure, on constate que la fréquence maximale effectivement atteignable avec les paramètres de la table 6.3 est, en accord avec les simulations de la section 4.3.1, légèrement inférieure à f_2 et de l'ordre de 13.6 kHz. Les composantes fréquentielles de fréquence supérieure à $f = 13.6$ kHz ne sont donc pas observables ici.

Différents types de mesures, schématiquement représentés sur la figure 6.4, ont été effectués :

- *Mesures selon z* : les vitesses normales induites dans le panneau par l'excitateur 6 (au centre du LaMAP de droite) ont été mesurées en 101 points espacés de 0.9 cm et placés sur un segment vertical de 90 cm passant par l'excitateur 6. Il a été fait de même pour l'excitateur 10 situé à proximité du bord gauche du LaMAPs de droite. Ces deux excitateurs ont été choisis pour observer l'importance de la position de l'excitateur sur le panneau sur sa réponse en vitesse.
- *Mesures selon (Oy)* : les vitesses normales induites dans le panneau de droite par chacun des 24 excitateurs ont été mesurées en 69 points espacés de 4 cm et placés sur le segment horizontal reliant les 12 excitateurs du panneau de droite.
- *Mesures autour d'un excitateur* : les vitesses normales induites dans le panneau par l'excitateur 6 (au centre du LaMAP de droite) ont été mesurées en 121 points répartis sur une grille de 11 points par 11 points ayant un espacement de 2 cm et centrée sur l'excitateur 6. Il a été fait de même pour l'excitateur

Paramètre	Valeur
f_1	20 Hz
f_2	15 kHz
f_s	44.1 kHz
T	10 s
N	4

TAB. 6.3 – Paramètres utilisés pour l'obtention des réponses impulsionnelles en vitesse des panneaux par la méthode proposée en section 4.2.

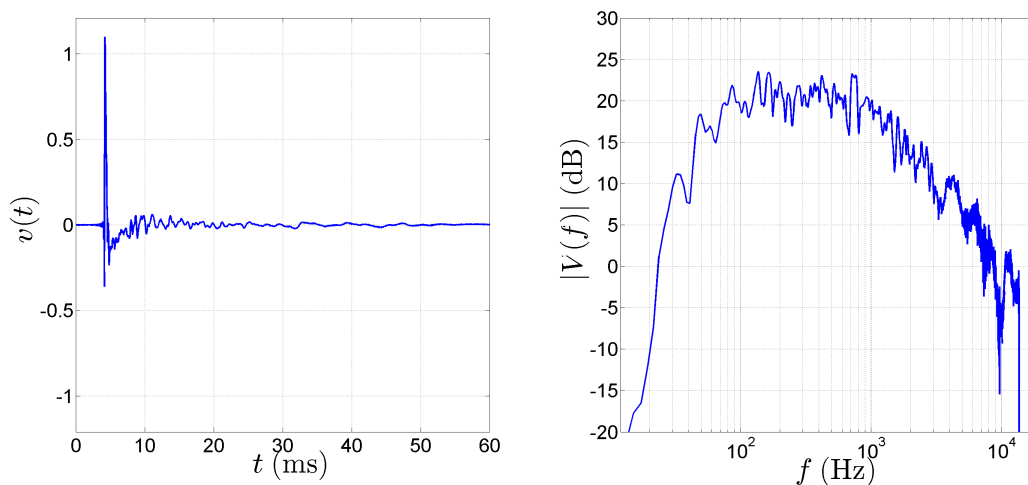


FIG. 6.3 – Mesure réalisée à proximité de l'excitateur 10 ($z = -4.5$ cm).

Gauche : réponse impulsionnelle $v(t)$ en fonction du temps. *Droite* : amplitude de la transformée de Fourier $V(f)$ de la réponse impulsionnelle en fonction de la fréquence.

15 situé à droite du LaMAP de gauche. Ces deux excitateurs ont été choisis pour juger d'éventuelles dissemblances existant entre les deux LaMAPs.

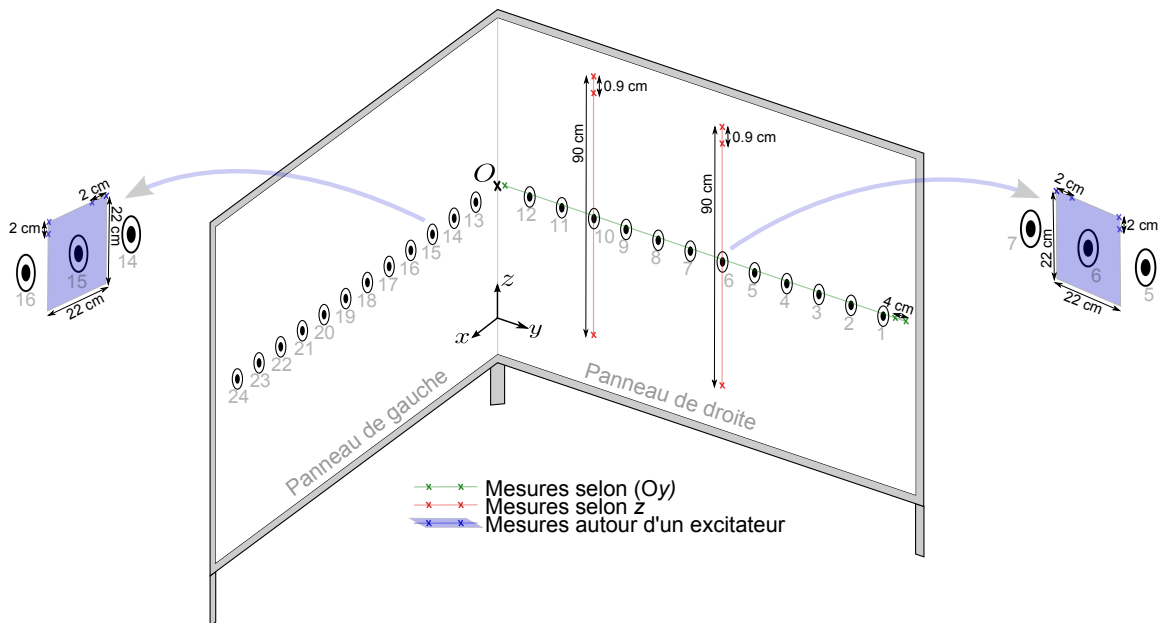


FIG. 6.4 – Schéma récapitulatif des différents types de mesures effectuées sur le dispositif.

6.3.2 Analyse temporelle

Les résultats expérimentaux sont d'abord analysés temporellement dans cette section.

Mesures selon z

Sur la figure 6.5, les réponses impulsionnelles $|v(z, t)|$ obtenues lors des mesures selon z pour l'excitateur 6 sont représentées dans l'espace et le temps. Ces réponses impulsionnelles ont été normalisées par la plus grande valeur spatio-temporelle obtenue sur la ligne. L'étendue de l'échelle selon z ($[-110 \text{ cm}, 90 \text{ cm}]$) correspond aux dimensions réelles du LaMAP sur lequel ont été effectuées les mesures. Les résultats obtenus pour l'excitateur 10 sont quasiment identiques à ceux de la figure 6.5 et ne sont donc pas présentés ici. Au moins pour ces deux excitateurs, leur position sur le panneau ne semble pas avoir de conséquence sur $|v(z, t)|$.

Propagation du front d'onde et corrélation avec le modèle théorique : l'impulsion part de $z \simeq 0 \text{ cm}$ à $t \simeq 0 \text{ ms}$ et se propage ensuite vers les z positifs et négatifs à mesure que t augmente. Apparaissent ensuite plusieurs réflexions qui viennent troubler le motif spatio-temporel régulier visible jusque là : le front d'onde rencontre d'abord le bord supérieur (première réflexion arrivant à $t = 2.32 \text{ ms}$), puis le bord inférieur (seconde réflexion arrivant à $t = 4.8 \text{ ms}$), et enfin les bords latéraux (réflexions non visibles).

Dans les plaques minces, plus la fréquence est élevée, plus les ondes se propagent vite. À z donné, ce sont donc les ondes de plus haute fréquence qui arrivent les premières. La vitesse de phase v_ϕ d'une onde de flexion de pulsation ω est déterminable à partir de l'équation de dispersion (6.3) :

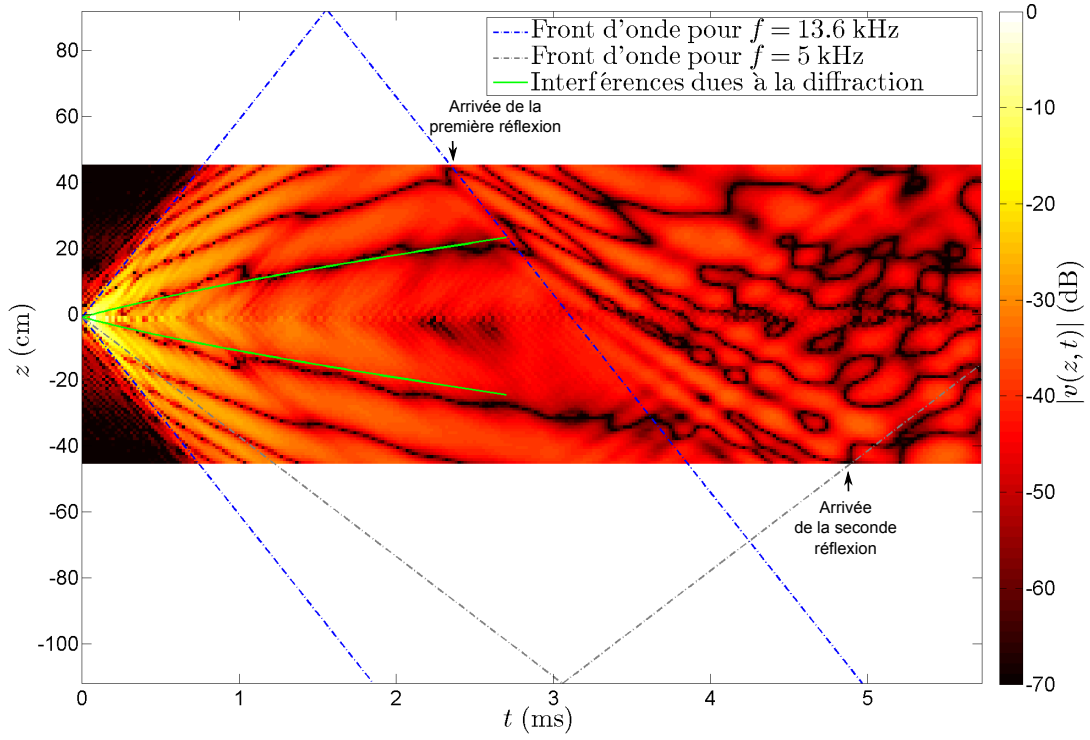


FIG. 6.5 – Représentation spatio-temporelle des réponses impulsionnelles $|v(z, t)|$ obtenues lors des mesures selon z à partir de l’excitateur 6.

Réponses impulsionnelles normalisées par la plus grande valeur spatio-temporelle obtenue selon z . L’étendue de l’échelle selon z correspond aux dimensions réelles du LaMAP sur lequel ont été effectuées les mesures ($[-110 \text{ cm}, 90 \text{ cm}]$). L’amplitude des réponses impulsionnelles en dB est indiquée par la couleur. Les résultats obtenus pour l’excitateur 10 sont quasiment identiques à ceux présentés ici.

$$v_\phi(\omega) = \frac{\omega}{k} = \overline{D}_0^{-1/4} \omega^{1/2} \quad (6.10)$$

Le trajet spatio-temporel d’un front d’onde se propageant dans le panneau peut ainsi être prédit en fonction de sa pulsation :

$$z(\omega, t) = v_\phi(\omega)t = \overline{D}_0^{-1/4} \omega^{1/2} t \quad (6.11)$$

Comme souligné à la section 6.3.1, la fréquence maximale observable avec le protocole expérimental choisi est de 13.6 kHz. Sur la figure 6.5 sont tracés les trajets des fronts d’onde issus de $z \simeq 0 \text{ cm}$ à $t \simeq 0 \text{ ms}$ correspondant à $f = 13.6 \text{ kHz}$. On constate que les fronts d’onde théoriques délimitent parfaitement les fronts d’ondes observés expérimentalement. Le modèle de la section 6.2 semble donc très bien prédire la vitesse de propagation des ondes dans le panneau².

2. Sur la figure 6.11, on constate toutefois qu’une part marginale de l’énergie précède temporellement les fronts d’onde prédits pour $f = 13.6 \text{ kHz}$. Cette part ne semble cependant pas être due à la propagation d’ondes de flexion. Il est probable qu’elle soit un artefact de la méthode de mesure causé par le fait que le système vibro-acoustique étudié ne soit pas tout à fait modélisable par un système de Hammerstein en cascade (voir section 4.4 et [Torrás-Rosell & Jacobsen 2011]).

Réflexions successives et conditions aux limites : les positions z des réflexions étant connues, l'étude de leurs instants d'arrivée renseigne au travers de l'équation (6.11) sur la fréquence maximale des ondes de flexion composant cette réflexion. Si le temps d'arrivée de la première réflexion est analysé, alors on trouve que la fréquence la plus haute contenue dans cette réflexion est de l'ordre de 13.5 kHz. Les conditions aux limites sur la partie haute de ce LaMAP renvoient donc les hautes fréquences. Si l'on procède à la même analyse sur la seconde réflexion, on trouve alors que la fréquence la plus haute contenue dans cette réflexion est de l'ordre de 5 kHz. Les conditions aux limites sur la partie basse de ce LaMAP ne renvoient donc visiblement pas les hautes fréquences. Cette asymétrie entre conditions aux limites en $z = 90$ cm et $z = -110$ cm peut s'expliquer par le fait que le LaMAP est simplement glissé dans un cadre métallique rainuré et garni de mousse. La partie basse du panneau est en contact forcé avec la mousse et l'armature par action du poids du panneau et des excitateurs et la partie haute du panneau est laissée quasiment libre, d'où l'asymétrie observée des conditions aux limites.

Interférences créées par la diffraction par les excitateurs voisins : le dernier point attirant l'attention sur la figure 6.5 est la présence de raies noires, telles que $v(z, t) \simeq 0$, réparties de façon quasiment parallèles les unes aux autres dans le plan (t, z) . Le modèle théorique de la section 6.2 ne prédit pas l'existence de ces raies. Il est probable que ces raies noires soient dues à des interférences entre le champ direct généré par l'excitateur 6 et les réflexions de ce champ sur les excitateurs voisins. On s'intéresse uniquement à la première frange d'interférence due au champ généré par les voisins directs de l'excitateur 6. Supposons que les excitateurs agissent comme des réflecteurs parfaits de coefficient de réflexion en vitesse égal à -1 . Les ondes émises par l'excitateur 6 et reçues sur les excitateurs 5 et 7 sont en phase si la longueur d'onde associée λ est égale à l'écart inter-excitateur : $\lambda = d_{\text{exc}} = 20$ cm. Selon l'hypothèse faite sur le coefficient de réflexion, les ondes réfléchies par les excitateurs 5 et 7 sont donc en opposition de phase si $\lambda = d_{\text{exc}}$, c'est à dire pour $f_{\text{exc}} = \omega_{\text{exc}}/(2\pi) = 660$ Hz. Alors, la première frange d'interférences dans le plan (t, z) correspond aux points (t, z) où la différence de phase entre les deux ondes est égale à la différence de phase engendrée par la différence de chemin parcourue par les ondes directes et réfléchies (voir figure 6.6) :

$$\omega t - kz_{\text{int}}(t) = -k\sqrt{z_{\text{int}}^2(t) + d_{\text{exc}}^2} - kd_{\text{exc}} \quad (6.12)$$

On obtient ainsi pour la première frange d'interférence générée par les voisins directs de l'excitateur 6 :

$$z_{\text{int}}(t) = \frac{t}{2} v_{\phi}(\omega_{\text{exc}}) \frac{2 + \frac{tv_{\phi}(\omega_{\text{exc}})}{d_{\text{exc}}}}{1 + \frac{tv_{\phi}(\omega_{\text{exc}})}{d_{\text{exc}}}} \quad (6.13)$$

Ce lieu est tracé sur la figure 6.5 et l'on constate qu'il correspond avec une bonne précision à la position de la première raie noire observée expérimentalement, malgré la simplicité du modèle utilisé et l'approximation que les excitateurs agissent comme des réflecteurs parfaits de coefficient de réflexion en vitesse égal à -1 .

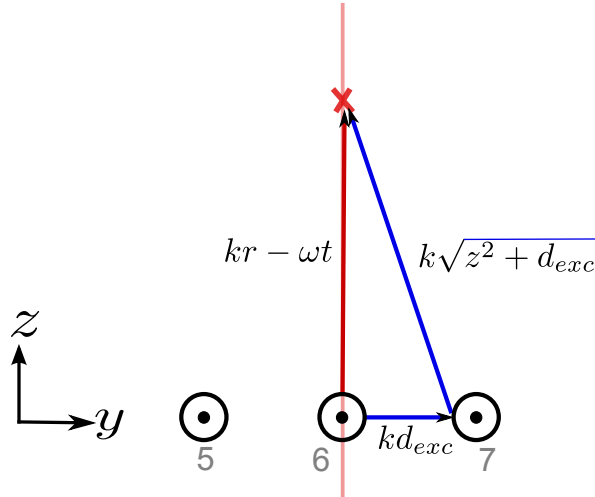


FIG. 6.6 – Différence de phase engendrée par la différence de chemin parcouru par les ondes directes et réfléchies.

Mesures selon (Oy)

Sur la figure 6.7, les réponses impulsionnelles $|v(y, t)|$ obtenues lors des mesures selon (Oy) pour l'excitateur 3 sont représentées dans l'espace et le temps. Ces réponses impulsionnelles ont été normalisées par la plus grande valeur spatio-temporelle obtenue sur la ligne. L'étendue de l'échelle verticale ([0 cm, 260 cm]) correspond aux dimensions réelles du LaMAP sur lequel ont été effectuées les mesures. L'amplitude des réponses impulsionnelles en dB est indiquée par la couleur. Les résultats obtenus pour les autres excitateurs sont qualitativement très similaires à ceux de la figure 6.5 et ne sont donc pas présentés ici. La position de l'excitateur sur le LaMAP ne semble donc pas influencer qualitativement le comportement des différents excitateurs.

Propagation du front d'onde et corrélation avec le modèle théorique : la plus faible résolution spatiale des mesures selon (Oy) (4 cm au lieu de 0.9 cm) conduit à des résultats moins facilement interprétables. Sur la figure 6.7 sont tracés les trajets des fronts d'onde issus de $y \simeq 210$ cm (position de l'excitateur 3) à $t \simeq 0$ ms correspondant à $f = 13.6$ kHz. On constate que comme précédemment ces fronts d'onde théoriques délimitent correctement les fronts d'ondes observés expérimentalement. Le modèle de la section 6.2 semble donc là encore bien prédire la vitesse de propagation des ondes dans le panneau.

Réflexions successives et conditions aux limites : la première réflexion, ayant lieu à $t = 1$ ms peut être étudiée en se basant sur l'équation (6.11). Le trajet correspondant à des ondes de fréquence égale à 13.5 kHz issues de cette réflexion semble visuellement délimiter correctement la zone où le motif spatio-temporel régulier se trouble. Les conditions aux limites sur la partie droite de ce LaMAP semblent donc renvoyer les hautes fréquences. Une analyse similaire menée sur les données des excitateurs situés près du bord gauche donne qualitativement les mêmes résultats. Les conditions aux limites des bords gauche et droit du LaMAP étudié semblent donc similaires à la condition aux limites présente sur la partie haute du LaMAP.

Interférences créées par la diffraction par les excitateurs voisins : comme sur la figure 6.5, on note aussi sur la figure 6.7 la présence de raies noires, telles que $v(y, t) \simeq 0$, réparties de façon quasiment parallèles les unes aux autres dans le plan (t, y) . Il est probable que ces raies noires soient aussi dues

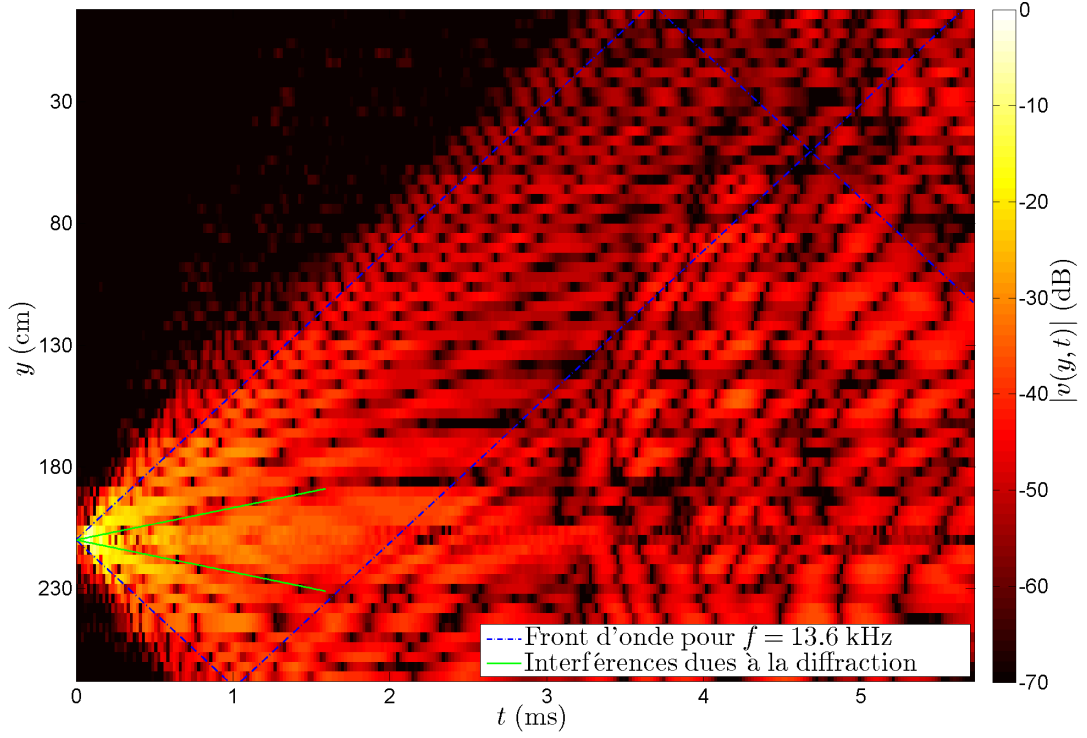


FIG. 6.7 – Représentation spatio-temporelle des réponses impulsionnelles $v(y, t)$ obtenues lors des mesures selon (Oy) pour l'excitateur 3.

Réponses impulsionnelles normalisées par la plus grande valeur spatio-temporelle obtenue sur la ligne. L'étendue de l'échelle verticale ([0 cm, 260 cm]) correspond aux dimensions réelles du LaMAP sur lequel ont été effectuées les mesures. L'amplitude des réponses impulsionnelles en dB est indiquée par la couleur. Les résultats obtenus pour les autres excitateurs sont qualitativement similaires à ceux présentés ici.

à des interférences entre le champ direct généré par l'excitateur 3 et les réflexions de ce champ sur les excitateurs voisins. Comme précédemment, on s'intéresse uniquement à la première frange d'interférence due au champ généré par les voisins directs de l'excitateur 3. On suppose toujours que les excitateurs agissent comme des réflecteurs parfaits de coefficient de réflexion en vitesse égal à -1 . Les ondes émises par l'excitateur 3 et reçues sur les excitateurs 2 et 4 sont en phase si la longueur d'onde associée λ est égale à l'écart inter-excitateur : $\lambda = \Delta_{\text{exc}} = 20$ cm. Selon l'hypothèse faite sur le coefficient de réflexion, les ondes réfléchies par les excitateurs 2 et 4 sont donc en opposition de phase pour $\lambda = \Delta_{\text{exc}}$, c'est à dire pour $f_{\text{exc}} = 660$ Hz. Alors, la première frange d'interférences dans le plan (t, y) correspond au trajet d'une onde de fréquence f_{exc} se propageant depuis l'excitateur :

$$y_{\text{int}}(t) = v_{\phi}(\omega_{\text{exc}})t \quad (6.14)$$

Ce lieu est tracé sur la figure 6.7 et l'on constate qu'il correspond avec une bonne précision à la position de la première raie noire observée expérimentalement pour $t < 1.5$ ms. Pour $t > 1.5$ ms, les réflexions provenant des autres voisins contribuent aussi et ce modèle ne prenant en compte que les voisins directs ne semble pas suffire.

Influence du LaMAP de gauche sur le LaMAP de droite

Les vitesses $|v(y, t)|$ engendrées sur le LaMAP de droite par les excitateurs fixés sur le LaMAP de gauche (*i.e.* ceux numérotés de 13 à 24 sur la figure 6.4) ayant aussi été mesurées, il est possible d'étudier l'influence du LaMAP de gauche sur le LaMAP de droite. Sur la figure 6.8 sont tracées les représentations spatio-temporelles des réponses impulsionnelles $|v(y, t)|$ engendrées sur le LaMAP de droite par les excitateurs 13 et 16. Ces réponses impulsionnelles ont été normalisées par la plus grande valeur spatio-temporelle obtenue sur le LaMAP de droite parmi tous les excitateurs (ceux du panneau de gauche et de droite confondus). L'étendue de l'échelle verticale ([0 cm, 260 cm]) correspond aux dimensions réelles du LaMAP sur lequel ont été effectuées les mesures. L'amplitude des réponses impulsionnelles en dB est indiquée par la couleur. Les résultats obtenus pour les autres excitateurs du LaMAP de gauche sont qualitativement similaires à ceux présentés ici.

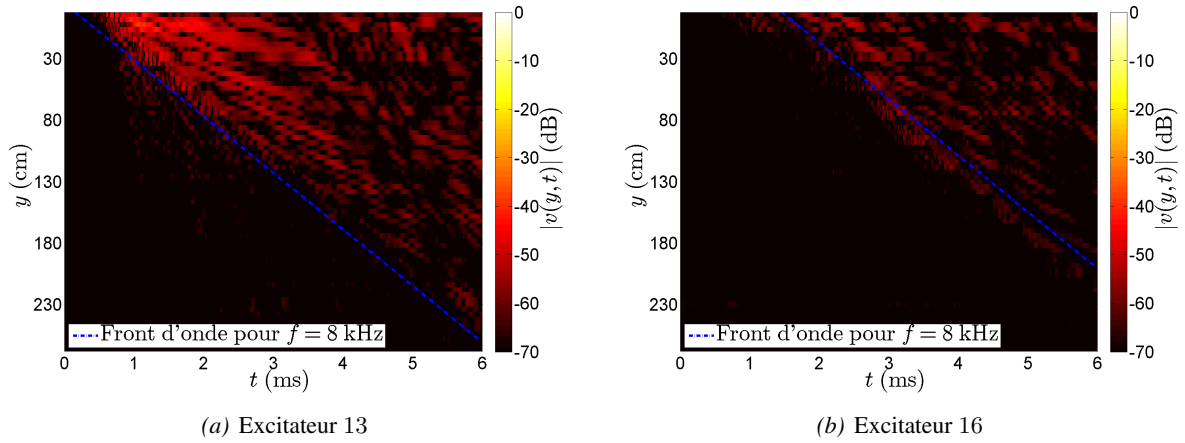


FIG. 6.8 – Représentation spatio-temporelle des réponses impulsionnelles $|v(y, t)|$ engendrées sur le LaMAP de droite par les excitateurs 13 et 16, fixés sur le LaMAP de gauche.

Réponses impulsionnelles normalisées par la plus grande valeur spatio-temporelle obtenue sur le LaMAP de droite parmi tous les excitateurs. L'étendue de l'échelle verticale ([0 cm, 260 cm]) correspond aux dimensions réelles du LaMAP sur lequel ont été effectuées les mesures. L'amplitude des réponses impulsionnelles en dB est indiquée par la couleur. Les résultats obtenus pour les autres excitateurs du LaMAP de gauche sont qualitativement similaires à ceux présentés ici.

De façon générale, la figure 6.8 indique que l'influence du LaMAP de gauche sur le LaMAP de droite est assez faible avec un maximum de $\simeq -38$ dB pour l'excitateur 13 et de $\simeq -50$ dB pour l'excitateur 16. La valeur de référence est le maximum, pour tous les excitateurs du LaMAP de droite, de la vitesse mesurée sur le panneau de droite.

Deux types d'interactions entre les deux LaMAPs sont possibles : des interactions acoustiques, qui seraient dues aux réflexions du rayonnement du LaMAP de gauche sur le LaMAP de droite, et des interactions mécaniques, qui seraient dues à une transmission mécanique³ entre les deux LaMAPs au niveau de la jonction centrale du SMART-I² (voir figure 6.4). Le fait que sur la figure 6.8 la propagation des ondes réfléchies

3. Il est probable qu'il n'y ait que peu d'énergie transmise d'un panneau à l'autre par l'intermédiaire de la structure en aluminium encadrant le SMART-I² étant donné que les LaMAPs sont en contact avec cette structure uniquement par l'intermédiaire de bandes de mousse absorbante.

se fasse suivant un cône rappelle cependant les figures 6.5 et 6.7. Cela laisse donc penser que l'influence du LaMAP de gauche sur le LaMAP de droite est plutôt due à une transmission mécanique d'énergie via la jonction centrale du SMART-I². Les tracés des trajets dans le plan (t, y) des fronts d'onde se propageant à $f = 8$ kHz depuis les excitateurs 13 et 16 au travers du LaMAP de droite semblent corroborer ce point. La jonction laisserait donc passer plutôt les fréquences inférieures à 8 kHz.

Mesures autour des excitateurs 6 et 15

Sur la figure 6.9 sont tracées dans le domaine spatio-temporel les vitesses $|v(y, z, t)|$ mesurées lorsque l'excitateur 15 est actif. Cette figure présente des prises de vue figées dans le temps à six instants consécutifs : $t = 0$ ms, $t = 0.02$ ms, $t = 0.05$ ms, $t = 0.35$ ms, $t = 3$ ms et $t = 3.5$ ms. Les valeurs mesurées ont été normalisées par la plus grande valeur spatio-temporelle obtenue sur la grille de mesure. Les échelles verticale et horizontale couvrent seulement 20 cm du LaMAP sur lequel ont été effectuées les mesures. L'amplitude des réponses impulsionnelles en dB est indiquée par la couleur. Les résultats obtenus pour l'excitateur 6 sont qualitativement similaires à ceux montrés ici et ne sont donc pas présentés.

En premier lieu, la figure 6.9 permet de discuter l'hypothèse émise à la section 6.2.3 selon laquelle les excitateurs auraient une action ponctuelle sur les LaMAPs. Sur cette figure sont indiquées la position du centre de l'excitateur et celle de l'anneau sur lequel sont réparties les zones de contact entre l'excitateur et le LaMAP (voir section 6.2.3). À l'instant $t = 0$ ms, on constate que l'action de l'excitateur sur le panneau est bien répartie sur 3 points distribués sur le pourtour de la zone de contact entre l'excitateur et le panneau. Cela est cohérent avec le fait que l'excitateur n'est effectivement collé au LaMAP qu'en 3 points de la zone de contact (voir figure 6.1). D'autre part, à $t = 0.02$ ms, les contributions de chacune de ces zones ont fusionné pour donner naissance à un profil de vitesse ayant un diamètre similaire à celui de l'excitateur. À $t = 0.05$ ms, ce profil commence à se propager et donc à s'élargir. À $t = 0.35$ ms, l'excitateur va vers sa position de repos et l'impulsion qu'il a émise se propage dans le LaMAP. Enfin, à $t = 3$ ms et $t = 3.5$ ms, des fronts d'onde issus des réflexions sur les voisins directs reviennent vers l'excitateur 15. Ainsi, pour $t \geq 0.02$ ms, c'est à dire une fois que les *hautes fréquences* se sont propagées, l'hypothèse selon laquelle l'excitateur a une action ponctuelle sur le panneau semble raisonnable. Cela est cohérent avec l'analyse fréquentielle effectuée à la section 6.9.

Enfin, l'analyse de la figure 6.9 permet aussi de discuter des effets de l'anisotropie du matériau constitutif des LaMAPs. En effet, à $t = 0.35$ ms, on peut voir le front d'onde résultant d'une impulsion émise par un LaMAP. L'impulsion initiale semble se distribuer équitablement entre toutes les directions du LaMAP : le front d'onde observé est quasiment circulaire. Il semble ainsi aussi raisonnable de faire l'hypothèse, en première approche, que le LaMAP est constitué d'un matériau isotrope.

6.3.3 Analyse fréquentielle

Dans cette section, les résultats expérimentaux sont analysés fréquemment et comparés aux prédictions faites par le modèle simple présenté à la section 6.2.3 (plaque mince, homogène, isotrope, infinie, et excitée ponctuellement).

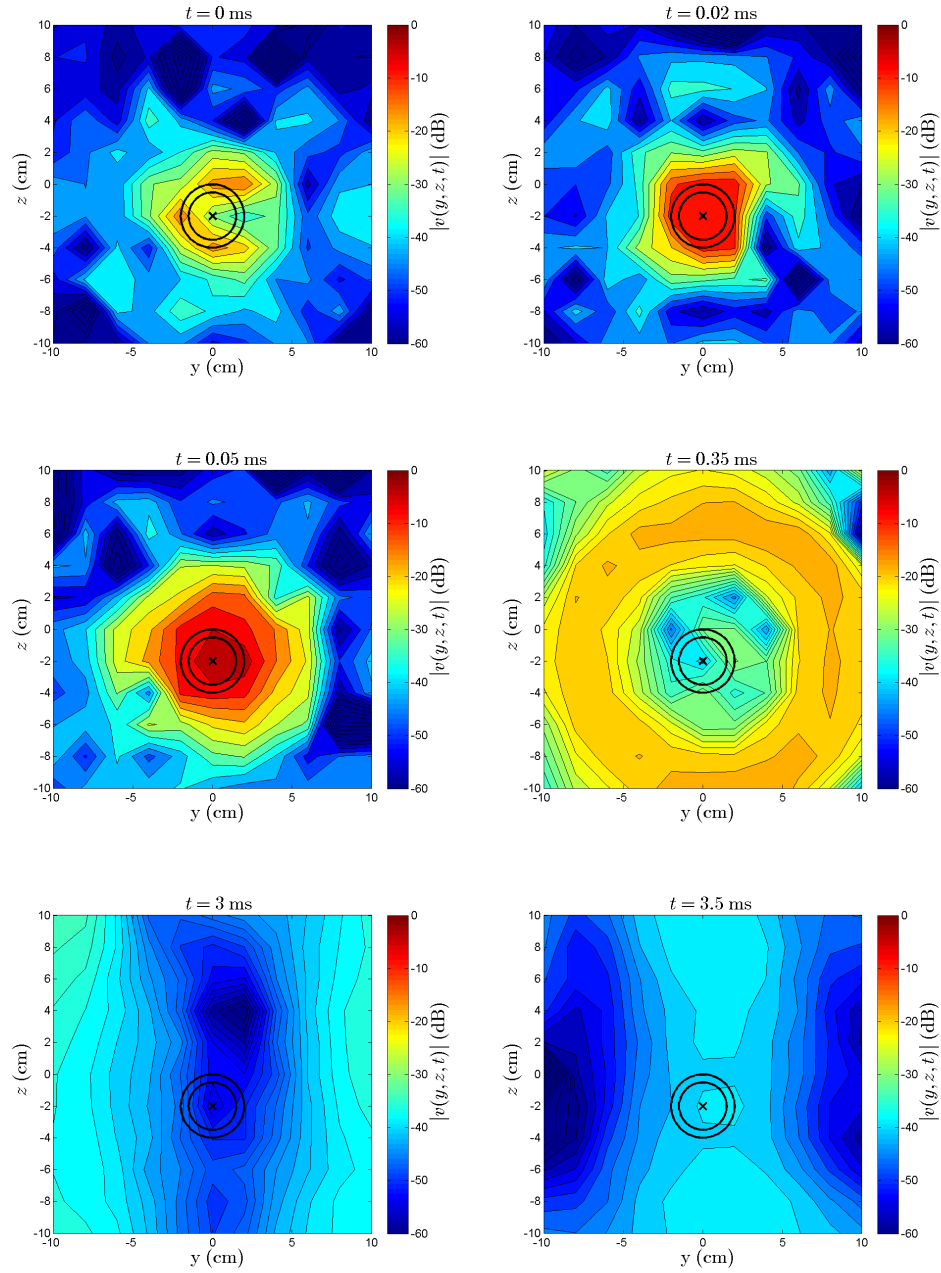


FIG. 6.9 – Propagation de la réponse impulsionnelle en vitesse $|v(y, z, t)|$ autour de l'excitateur 15.

Prises de vue figées dans le temps à six instants consécutifs $t = 0$ ms, $t = 0.02$ ms, $t = 0.05$ ms, $t = 0.35$ ms, $t = 3$ ms et $t = 3.5$ ms. Les valeurs mesurées ont été normalisées par la plus grande valeur spatio-temporelle obtenue sur la grille de mesure. Les échelles verticale et horizontale couvrent seulement 20 cm du LaMAP sur lequel ont été effectuées les mesures. L'amplitude des réponses impulsionnelles en dB est indiquée par la couleur. Les résultats obtenus pour l'excitateur 6 sont qualitativement similaires à ceux présentés ici.

Mesures selon z

La figure 6.10 représente l'amplitude et la phase des profils de vitesse $V(k, z)$ expérimentaux et théoriques, obtenus pour les fréquences $f = 180$ Hz, $f = 800$ Hz et $f = 1500$ Hz pour l'excitateur 6. Les profils expérimentaux sont normalisés de façon à avoir une amplitude unité et une phase nulle à la position de l'excitateur, *i.e.* à $z = 0$ cm. Les profils théoriques sont obtenus en utilisant les équations (6.2) et (6.9) et en prenant $r = |z|$. À titre indicatif, les longueurs d'onde λ données par l'équation (6.9) sont aussi mentionnées pour chacune des fréquences étudiées.

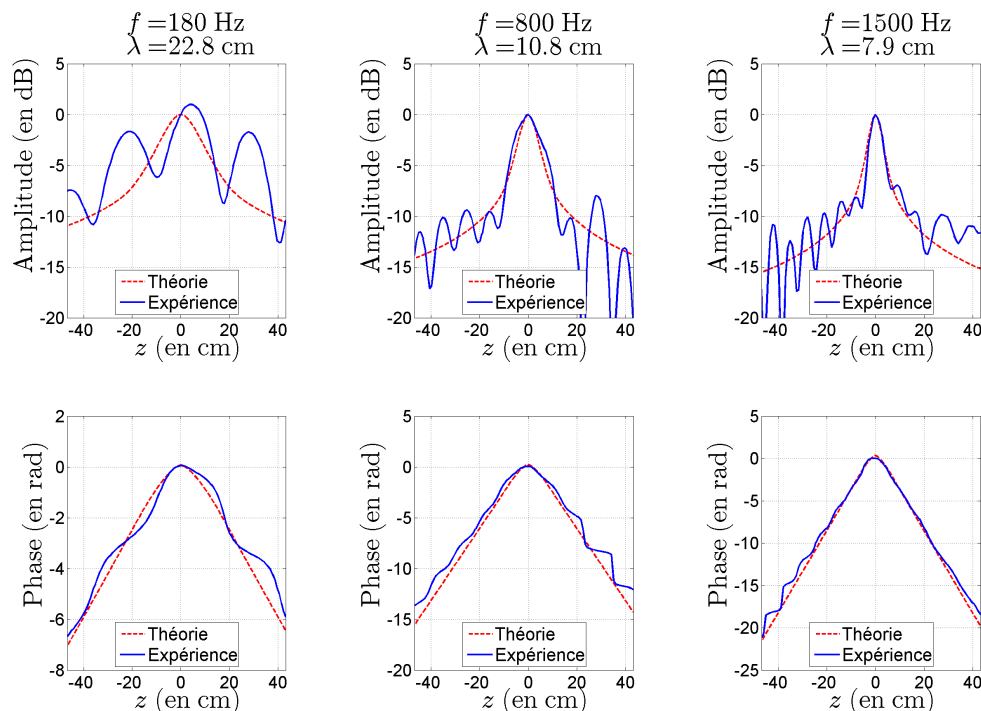


FIG. 6.10 – Amplitude et phase des profils de vitesse $V(k, z)$ expérimentaux et théoriques, obtenus pour les fréquences $f = 180$ Hz, $f = 800$ Hz et $f = 1500$ Hz pour l'excitateur 6.

Les profils expérimentaux sont normalisés de façon à avoir une amplitude unité et une phase nulle à la position de l'excitateur, *i.e.* à $z = 0$ cm. Les profils théoriques sont obtenus en utilisant les équations (6.2) et (6.9) et en prenant $r = |z|$. À titre indicatif, les longueurs d'onde λ données par l'équation (6.9) sont aussi mentionnées pour chacune des fréquences étudiées.

Sur la figure 6.10, on constate que les résultats théoriques et expérimentaux sont qualitativement en assez bon accord pour les trois fréquences étudiées, que ce soit concernant l'amplitude ou la phase des profils de vitesse $V(k, z)$. Une des différences majeures observée entre la théorie et l'expérience réside dans la présence de “zéros” en amplitude, associés à des sauts de phase de π , qui apparaissent pour certaines valeurs de z et pour toutes les fréquences étudiées. Il s'avère aussi que les distances entre deux “zéros” consécutifs correspondent assez bien aux longueurs d'onde λ obtenues théoriquement. Cela signifie donc que les conditions aux limites restent influentes assez haut en fréquence et qu'il ne sera pas possible de considérer la plaque comme infinie. Le comportement modal de la plaque est particulièrement bien illustré pour $f = 180$ Hz. En effet, pour cette fréquence, le maximum d'amplitude observé expérimentalement ne

correspond pas au point d'excitation : un mode possédant un ventre à $z \simeq 4$ cm est excité et sa déformée modale est observée. On constate aussi une asymétrie des profils de vitesse entre $z > 0$ et $z < 0$. Cette asymétrie peut avoir deux explications : elle peut soit provenir du fait que les excitateurs ne sont pas situés au milieu des LaMAPs, soit provenir des conditions aux limites hautes et basses qui sont différentes (voir section 6.3.2).

Mesures selon (Oy)

La figure 6.11 représente l'amplitude et la phase des profils de vitesse $V(k, y)$ expérimentaux et théoriques, obtenus pour les fréquences $f = 180$ Hz, $f = 400$ Hz et $f = 800$ Hz pour l'excitateur 8. Les profils expérimentaux sont normalisés de façon à avoir une amplitude unité et une phase nulle à la position de l'excitateur 8, *i.e.* à $y = 0$ cm. Les profils théoriques sont obtenus en utilisant les équations (6.2) et (6.9) et en prenant $r = |y|$. À titre indicatif, les longueurs d'onde λ données par l'équation (6.9) sont aussi mentionnées pour chacune des fréquences étudiées.

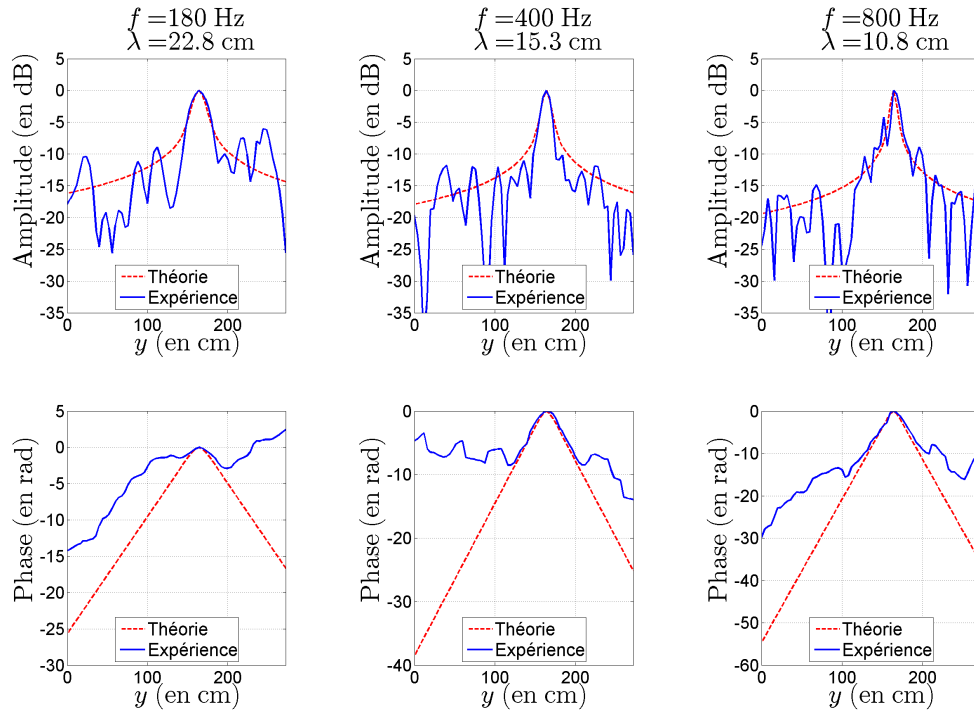


FIG. 6.11 – Amplitude et phase des profils de vitesse $V(k, y)$ expérimentaux et théoriques, obtenus pour les fréquences $f = 180$ Hz, $f = 400$ Hz et $f = 800$ Hz pour l'excitateur 8.

Les profils expérimentaux sont normalisés de façon à avoir une amplitude unité et une phase nulle à la position de l'excitateur. Les profils théoriques sont obtenus en utilisant les équations (6.2) et (6.9) et en prenant $r = |y|$. À titre indicatif, les longueurs d'onde λ données par l'équation (6.9) sont aussi mentionnées pour chacune des fréquences étudiées

Sur la figure 6.11, on observe que l'adéquation entre théorie et expérience est bien moins bonne que sur la figure 6.10. En effet, même si l'amplitude décroît expérimentalement de façon assez similaire à ce que prédit la théorie, ce n'est pas du tout le cas pour la phase, quelque soit la fréquence étudiée. Là

encore apparaissent des “zéros” en amplitude, associés à des sauts de phase de π , pour certaines valeurs de y et pour toutes les fréquences étudiées. Les distances entre deux “zéros” consécutifs ne correspondent pas forcément bien aux longueurs d’ondes λ obtenus théoriquement. La présence des autres excitateurs semble donc avoir une grande importance dans la dimension horizontale et le modèle simple proposé à la section 6.2.3 est dans ce cas mis en défaut dès que $|y|$ dépasse $\simeq 20$ cm : la distance inter-excitateurs.

6.4 Conclusion et perspectives

Ce chapitre présente une première description de différents phénomènes influençant le comportement dynamique du SMART-I² (conditions aux limites, diffractions par les excitateurs voisins, interactions ayant lieu entre les deux LaMAPs) en confrontant résultats expérimentaux et prédictions faites par un modèle simple. Les données expérimentales recueillies étant très riches, de nombreuses analyses sont encore à réaliser pour en tirer pleinement profit.

En ce qui concerne les conditions aux limites, cette *étude préliminaire* révèle que le LaMAP étudié semble ne pas posséder des conditions aux limites homogènes sur ses quatre côtés. Les conditions aux limites sont globalement réfléchissantes en haut, à gauche et à droite. La condition aux limites en bas semble absorber les hautes fréquences. Dans le domaine temporel, des interférences entre le champ direct et le champ réfléchi par les excitateurs voisins sont aussi observées. Enfin, l’influence du LaMAP de gauche sur le LaMAP de droite est étudiée : la jonction mécanique centrale du SMART-I² semble transmettre mécaniquement une part marginale de l’énergie d’un panneau à l’autre.

Cette étude expérimentale a aussi permis de juger de la capacité d’un modèle mécanique simple à prédire le comportement dynamique des LaMAPs : un modèle de plaque mince, isotrope, homogène et infinie excitée ponctuellement. De façon générale, ce modèle simple permet de bonnes prédictions qualitatives (vitesse de propagation dans la plaque, amplitude et phase dans la direction verticale). Cependant, même si le recouvrement modal est important dans le domaine de fréquence considéré ($\mu > 1$), les conditions aux limites interviennent quand même de manière forte sur la dynamique des panneaux : le modèle de plaque infinie ne suffit donc pas. Il semblerait donc approprié de chercher à modéliser les LaMAPs plutôt à l’aide d’un modèle statistique. Cependant, les approches statistiques habituelles (comme la SEA⁴ par exemple) sont énergétiques et prédisent les résultats *en moyenne*. Or pour l’application visée ici, *i.e.* l’holophonie, les détails du comportement dynamique nous intéressent aussi “*dans une certaine mesure*” (qu’il conviendra de déterminer). Il est donc important d’arriver à trouver un descripteur de la dynamique des LaMAPs qui soit adapté à l’objectif visé.

L’intérêt de disposer d’un modèle permettant de prédire précisément la dynamique des LaMAPs est considérable. En effet, la connaissance de la dynamique de ces panneaux permet d’envisager un contrôle fin du comportement vibratoire des LaMAPs. Le problème de la synthèse d’un champ sonore (WFS, voir section 2.5.3) pourrait ainsi être abordé, non plus d’un point de vue acoustique, mais d’un point de vue vibratoire. Un rendu sonore spatialisé pourrait alors être généré en synthétisant sur le panneau le champ de vitesse correspondant à une source sonore virtuelle placée à une position donnée dans l’espace. Cette procédure aurait de plus l’avantage d’être moins sensible à l’influence acoustique de la zone de restitution qu’une procédure uniquement basée sur un contrôle acoustique.

4. *Statistical Energy Analysis*, voir [Bruneau & Scelo 2006] par exemple.

Chapitre 7

Cohérence audio-visuelle de l'espace virtuel synthétisé par le SMART-I²

Sommaire

7.1	Introduction	175
7.2	Perception spatiale unimodale	176
7.2.1	Espace perceptif auditif	176
7.2.2	Espace perceptif visuel	182
7.3	Intégration spatio-temporelle audio-visuelle	184
7.3.1	Fenêtre spatio-temporelle d'intégration audio-visuelle	185
7.3.2	Facteurs cognitifs influents	188
7.3.3	Conséquences	189
7.4	Évaluation des performances offertes par le SMART-I²	190
7.4.1	Présentation générale	190
7.4.2	Analyse objective	192
7.4.3	Précision de localisation en azimut	194
7.4.4	Restitution de l'effet de parallaxe	198
7.4.5	Effet de l'apprentissage	202
7.5	Conclusion	203

7.1 Introduction

L'objectif de ce chapitre est de déterminer dans quelle mesure le SMART-I² permet à un utilisateur, plongé dans l'environnement de réalité virtuelle qu'il synthétise, d'utiliser ses capacités perceptives audio-visuelles et ses capacités motrices dans leur intégralité pour comprendre l'espace virtuel et se déplacer en son sein.

Dans ce but, la façon dont l'espace est perçu au travers des modalités auditive et visuelle est d'abord rappelée dans la section 7.2. Puis les conditions perceptives requises pour que les flux spatio-temporels d'informations provenant des modalités auditive et visuelle fusionnent et donnent naissance à un unique

flux audio-visuel sont décrites à la section 7.3. Deux expériences, ayant pour but d'estimer la précision de localisation en azimuth, d'étudier la capacité de restitution de l'effet de parallaxe auditive, ainsi que d'évaluer la cohérence perceptive du rendu audio-visuel proposé par le SMART-I² sont ensuite présentées à la section 7.4.

7.2 Perception spatiale unimodale

Dans cette section, il est supposé que le repère naturel dans lequel l'homme perçoit l'espace qui l'entoure est le repère de coordonnées sphériques (r, θ, ϕ) ayant pour origine l'intersection de l'axe interaural et de l'axe médian¹ (cf. figure 7.1). L'aptitude humaine à localiser visuellement et auditivement une source selon les trois coordonnées de ce repère égocentrique (distance, azimuth et élévation) est abordée dans cette section.

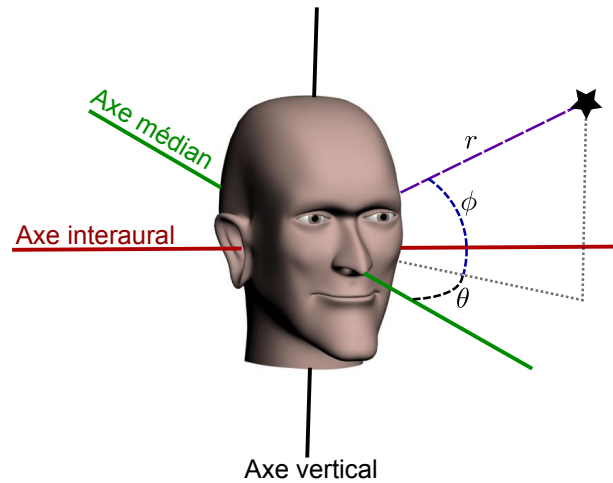


FIG. 7.1 – Repère sphérique associé à la tête du sujet. θ désigne l'azimut, ϕ l'élévation et r la distance de la source sonore par rapport à la tête. Le plan horizontal est défini comme le plan comprenant l'axe interaural et l'axe médian. Le plan médian est défini comme le plan comprenant l'axe vertical et l'axe médian.

7.2.1 Espace perceptif auditif

L'aptitude humaine à localiser auditivement une source selon les trois coordonnées du repère égocentrique décrit par la figure 7.1 est abordée dans cette section².

Estimation auditive de l'azimut

La localisation de la direction de sources sonores dans le plan horizontal (plan comprenant l'axe interaural et l'axe médian) est obtenue par la comparaison des signaux auditifs arrivant à chacune des oreilles.

1. En réalité, ceci n'est pas évident. En effet, le système visuel étant totalement tributaire de la position des yeux, il est plutôt oculo-centré. Le système auditif base quand à lui sa représentation de l'espace sur les deux capteurs statiques dont il dispose (les oreilles) et est plutôt auriculo-centré. L'hypothèse selon laquelle il existe un référentiel multimodal unique déterminé par la position relative des référentiels auditifs et visuels semble cependant valide [Hartnagel 2007].

2. Le fonctionnement de l'oreille n'est pas l'objet d'intérêt dans cette section et n'est par conséquent pas détaillé (voir par exemple [Fuchs et al. 2006a] pour plus d'informations à ce sujet).

Trois indices auditifs binauraux, repartis en deux catégories, servent à déterminer la direction de provenance (*i.e.* l'azimut, θ) des sources sonores dans le plan horizontal [Blauert 1999, Fuchs et al. 2006a] :

1. Les différences interaurales de temps (ITD)

- (a) *La différence de phase* entre les signaux sonores arrivant sur chacune des oreilles. Au-delà de 1500 Hz, les longueurs d'ondes associées aux ondes sonores sont inférieures au diamètre de la tête et donc les différences de phase ne sont plus informatives. L'oreille est donc sensible aux différences de phase surtout pour les fréquences inférieures à 1500 Hz.
- (b) *La différence de moment d'arrivée de l'enveloppe* du signal sonore au niveau de chaque oreille. Cet indice est surtout utilisé à partir de 1500 Hz, quand les différences de phase ne contiennent plus d'information.

2. Les différences interaurales de niveau sonore (ILD) dues à l'atténuation de l'onde sonore causée par la présence de la tête. Cet indice est valable sur toute la gamme de fréquences audibles, mais la diffraction des ondes sonores par la tête se manifeste plutôt à partir de 1500 Hz environ. En dessous de cette fréquence, les longueurs d'onde associées aux ondes sonores étant plus grandes que la tête, cette dernière n'est pas vue comme un obstacle et donc n'induit que peu de diffraction.

La figure 7.2 résume de façon simplifiée l'importance relative des différents indices binauraux utilisés par l'homme pour l'estimation de l'azimut d'une source sonore en fonction de la fréquence. La réalité est cependant bien plus complexe et les signaux reçus par l'homme ont beaucoup d'influence sur les indices binauraux qui seront utilisés pour les localiser. Pour des signaux large bande ayant de l'énergie à la fois au-dessous et au-dessus de 1.5 kHz, il semble par exemple que les différences interaurales de temps, et plus particulièrement les différences de phase, soient les indices les plus influents pour la localisation dans le plan azimutal [Wightman & Kistler 1992]. De façon générale, l'homme est en mesure de définir avec une précision de l'ordre de 3° (entre 1° et 4° en fonction du stimulus) la direction de provenance d'une source sonore dans le plan horizontal frontal à partir de ces trois indices, comme le montre par exemple la figure 7.4 pour $\theta = 0^\circ$ [Blauert 1999].

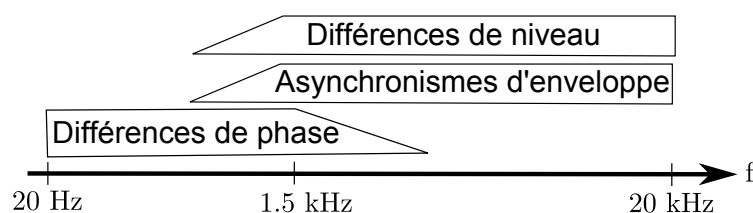


FIG. 7.2 – Schéma simplifié représentant l'importance relative des différents indices binauraux utilisés pour l'estimation de l'azimut d'une source sonore en fonction de la fréquence.

Estimation auditive de l'élévation

Dans le plan médian (*i.e.* le plan comprenant les axes médian et vertical) les champs sonores incidents sur chacune des oreilles étant égaux, les indices définis précédemment ne permettent pas d'estimer l'élévation ϕ des sources sonores. C'est alors le pavillon qui agit comme un filtre se comportant différemment en

fonction de la direction d'incidence des ondes sonores, le torse et les épaules de l'auditeur, par les réflexions qu'ils apportent, qui permettent l'estimation de l'élévation d'une source sonore située dans ce plan [Blauert 1999].

Pour une direction d'arrivée donnée (θ, ϕ) , un filtrage fréquentiel du signal émis par la source sonore est ainsi réalisé par les pavillons et le torse. Les fonctions de transfert entre une direction donnée (θ, ϕ) et les tympanes sont mesurables et appelées “*Head Related Transfer Functions*” (HRTFs) [Begault 1994, Blauert 1999, Nicol 2010]. C'est alors l'utilisation des indices binauraux décrits précédemment et l'apprentissage inconscient et individuel des filtrages fréquentiels réalisés par le pavillon et le torse en fonction de la direction d'incidence du son qui permettent à l'homme d'estimer la direction de provenance (θ, ϕ) d'une source sonore dans l'espace.

Il semble cependant que la capacité humaine d'estimation de l'élévation ϕ soit beaucoup moins précise que celle de l'azimut θ , et surtout qu'elle dépende assez fortement de la familiarité avec le contenu spectral de la source. Par exemple, la précision d'estimation de l'élévation d'une source située dans le plan médian varie de 4° pour un bruit blanc à 17° pour la voix d'une personne inconnue [Lokki & Grohn 2005, Godfroy et al. 2011]. Des résultats dans le plan médian et pour une voix familière sont présentés sur la figure 7.5.

Estimation auditive de la distance

Pour estimer la distance r entre une source sonore et lui, l'homme fait appel à différents indices acoustiques [Zahorik et al. 2005] :

1. **L'intensité** : le niveau sonore d'une source décroît au fur et à mesure que l'on s'éloigne. Ainsi, l'intensité est un indice très fin pour estimer les variations de la distance d'une source ou la distance séparant deux sources. Cependant l'intensité sonore ne permet pas, sans connaissance a priori de la source, d'estimer la distance absolue à laquelle se situe une source. C'est donc un indice de localisation relatif nécessitant seulement un traitement *bas-niveau*³ par le cerveau.
2. **Le rapport “champ réverbéré” sur “champ direct”** : dans une salle, au fur et à mesure que l'on s'éloigne d'une source, les réflexions multiples sur les parois de la salle des ondes issues de cette source (champ réverbéré) vont progressivement dominer, en amplitude, les ondes provenant directement de la source (champ direct). Le rapport entre l'énergie du *champ direct* et l'énergie du *champ réverbéré* reçus à la position d'écoute est ainsi un indice nécessitant un traitement *haut-niveau* par le cerveau et permettant l'estimation de la distance absolue à laquelle se situe une source sonore.
3. **Absorption de l'air** : lorsqu'un son se propage dans l'air, les hautes fréquences qu'il contient se trouvent atténuées. Cette atténuation constitue un indice de localisation permettant d'estimer la distance à laquelle une source sonore se trouve. Cependant, cette atténuation n'étant perceptible que pour des distances supérieures à 15 m, elle n'est utile que pour de grandes distances. Comme pour l'intensité, le spectre de la source doit être connu pour que cet indice soit informatif. Il s'agit donc plutôt d'un indice *haut-niveau* de localisation relatif.

3. Dans cette section ainsi que dans la section 7.2.2, la qualification d'indice de *bas niveau* est associée à tout indice dont le traitement par le cerveau est inné (par exemple réalisé à l'aide de neurones dédiés, comme pour l'ITD) et la qualification d'indice de *haut-niveau* est associée à toute indice acquis, *i.e.* résultant d'une procédure d'apprentissage. On pourra se référer à [Blauert 1999] pour plus de détails concernant ces dénominations.

4. **La familiarité avec le contenu spectral de la source** : pour des sources qui sont connues, leur contenu spectral renseigne sur leur proximité. Par exemple, à niveau sonore égal, la voix de quelqu'un qui chuchote paraîtra toujours plus proche que celle de quelqu'un parlant normalement. Cet indice *haut-niveau* permet alors d'estimer la distance d'une source sonore. Il est par contre important de noter qu'il ne constitue pas un indice *acoustique* à proprement parler.
5. **Indices binauraux** : dans une direction (θ, ϕ) donnée, et pour des sources situées assez loin de la tête de l'auditeur (*i.e.* > 1.5 m) les indices binauraux permettant l'estimation de l'azimut (ITD et ILD) sont quasiment indépendants de la distance. Cependant dans une direction (θ, ϕ) donnée et pour des sources assez proches de la tête de l'auditeur (*i.e.* < 1.5 m), les ITDs et ILDS vont varier avec la distance. Pour des distances inférieures à 1.5 m, les indices binauraux peuvent donc être utilisés pour estimer la distance absolue d'une source sonore. Ce sont des indices ne nécessitant qu'un traitement de *bas-niveau* par le cerveau.
6. **Indices dynamiques** : dans la vie quotidienne, les sources sonores et les auditeurs sont rarement immobiles. La parallaxe de mouvement, qui traduit le fait que pour un déplacement latéral donné de l'auditeur, des sources proches auront un déplacement relatif plus important que des sources lointaines permet d'estimer la distance absolue à laquelle se trouve une source sonore. Pour une source sonore convergeant vers l'auditeur (ou bien l'inverse), l'accroissement du niveau sonore perçu par l'auditeur permet aussi d'estimer le *temps avant impact* de la source et de l'auditeur, et donc indirectement, la distance absolue à laquelle se situe la source. Ainsi les variations dynamiques de l'intensité, qui est un indice relatif, permettent une estimation absolue de la distance.

L'importance des différents indices acoustiques absolus décrits précédemment est schématiquement représentée en fonction de la distance à laquelle se trouve la source sur la figure 7.3. Les performances humaines, en termes d'estimation de la distance à laquelle se trouve une source sonore, sont illustrées sur la figure 7.6.

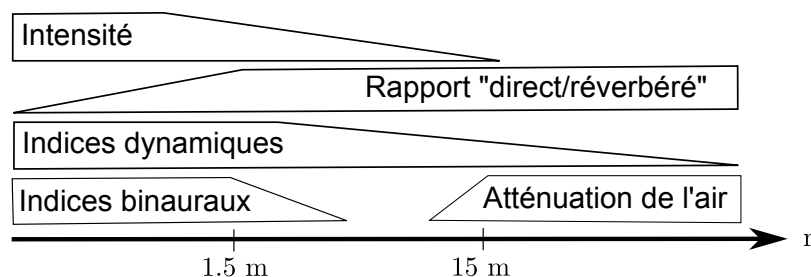


FIG. 7.3 – Schéma simplifié représentant l'importance relative des différents indices acoustiques utilisés pour l'estimation de la distance d'une source sonore en fonction de la distance à laquelle se trouve la source.

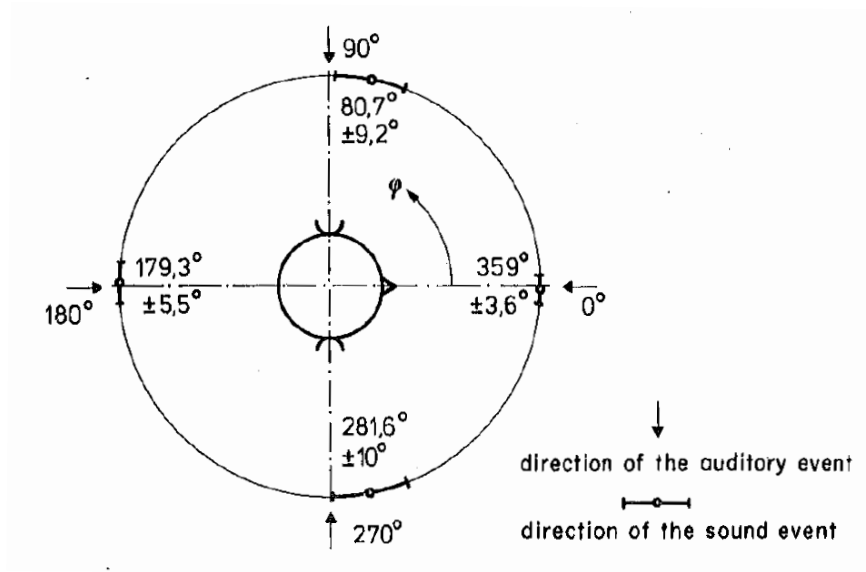


FIG. 7.4 – Récapitulatif des performances humaines en termes de localisation spatiale dans le plan horizontal (images tirée de [Blauert 1999], impulsions de bruit blanc de 100 ms).

Sur cette figure sont représentées les moyennes et écarts-types des réponses des sujets à des stimuli auditifs provenant de quatre directions différentes dans le plan azimutal : 0°, 90°, 180°, et 270°. Les directions de provenance des sons sont indiquées par les flèches, les réponses moyennes des sujets par les ronds blancs et les écarts-types par les portions d'arc de cercle plus épaisses.

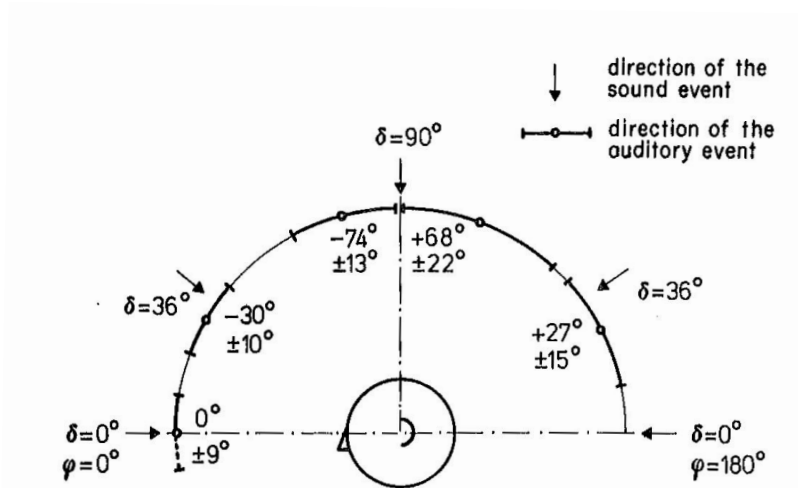


FIG. 7.5 – Récapitulatif des performances humaines en termes de localisation spatiale dans le plan médian (image tirée de [Blauert 1999], voix familière).

Sur cette figure sont représentées les moyennes et écarts-types des réponses des sujets à des stimuli auditifs provenant de quatre directions différentes dans le plan médian : 0° pour des azimuts de 0° et de 180°, 36° pour des azimuts de 0° et de 180°, 90°. Les directions de provenance des sons sont indiquées par les flèches, les réponses moyennes des sujets par les ronds blancs et les écarts-types par les portions d'arc de cercle plus épaisses.

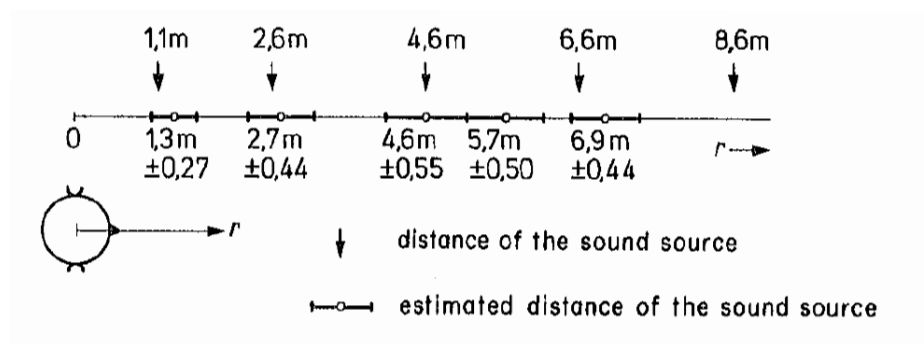


FIG. 7.6 – Récapitulatif des performances humaines en termes de localisation spatiale en distance (image tirée de [Blauert 1999], impulsion de bruit blanc).

Sur cette figure sont représentées les moyennes et écarts-types des réponses des sujets à des stimuli auditifs provenant de cinq distances différentes pour un azimut et une élévation nuls : 1.1 m, 2.6 m, 4.6 m, 6.6 m et 8.6 m,. Les distances des sources sont indiquées par les flèches, les réponses moyennes des sujets par les ronds blancs et les écarts-types par les portions de lignes plus épaisses.

7.2.2 Espace perceptif visuel

L'aptitude humaine à localiser visuellement une source selon les trois coordonnées du repère égocentrique décrit par la figure 7.1 est abordée dans cette section.

Physiologie de l'œil

L'œil est un organe très complexe, et nous n'allons ici préciser que quelques caractéristiques fondamentales de la perception du monde que nous offrent nos deux yeux. On pourra se référer à [Fuchs et al. 2006a] pour plus d'informations concernant ce sujet.

La fonction des yeux est de canaliser la lumière émise ou réfléchie par un objet, de façon à créer une image nette qui s'imprime sur la partie de l'œil recouverte de récepteurs sensoriels : *la rétine*. L'œil est constitué d'une succession de milieux transparents qui jouent le rôle d'une lentille convergente dont la focale globale peut varier par modification de la courbure du cristallin (voir figure 7.7).

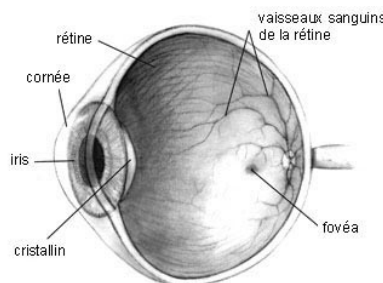


FIG. 7.7 – Vue en coupe de l'œil humain (extrait de [Fuchs et al. 2006a]).

La rétine est constituée de deux sortes de photo-récepteurs :

- *Les cônes* : concentrés sur la fovéa, ils sont sensibles à la longueur d'onde.
- *Les bâtonnets* : présents presque partout sur la rétine, ils sont très peu sensibles à la couleur et plutôt sensibles à l'intensité.

Même si les bâtonnets sont beaucoup plus nombreux que les cônes ($120 \cdot 10^6$ contre $6 \cdot 10^6$) ce sont les cônes qui contribuent en majorité à l'information transmise au cerveau. Par conséquent l'acuité visuelle n'est pas homogène sur tout le champ de vision.

Estimation visuelle de l'azimut et de l'élévation

Les performances d'un œil moyen pour l'estimation de l'azimut θ et de l'élévation ϕ sont les suivantes [Fuchs et al. 2006a] :

- L'acuité monoculaire est très grande dans un cône de 2° d'ouverture (champ de vision central) et se dégrade à mesure que l'objet visuel sort de ce cône (champ de vision périphérique). Cela est une conséquence directe de la répartition non-homogène des cônes sur la rétine. Dans ce cône, la valeur minimum de l'angle sous lequel deux points sont vus séparément est fonction du stimulus observé et est de l'ordre de $1'$ d'angle. Elle est de l'ordre de 2° pour un azimut de 25° (voir figure 7.9 et [Perrott 1993]).

- Le champ de vision *horizontal* est de 90° côté temple et 50° côté nez. Cela crée une zone de recouvrement d'environ 100° dans laquelle la vision est binoculaire. Le champ de vision *vertical* est de 45° vers le haut et de 70° vers le bas.

L'œil est donc capable d'une très forte précision, mais dans une zone très limitée de l'espace. Pour compenser cette limitation, il a la possibilité de bouger d'environ 15° dans les orbites à une vitesse maximale de 600°/s.

Estimation visuelle de la distance r

Contrairement à l'intuition, la perception de la distance ne découle pas uniquement de la vision binoculaire. En effet, de nombreux indices monoculaires servent aussi à la perception de la profondeur et donc de la troisième dimension spatiale. Les indices pour la perception de la distance sont assez nombreux et les principaux sont [Cutting 1997] :

1. **Hauteur dans le champ visuel** : la position du haut et du bas d'un objet par rapport à la ligne d'horizon permet de juger de la distance à laquelle se trouve un objet visuel. En effet, un objet situé très loin de l'observateur aura ses extrémités très proches de la ligne d'horizon tandis qu'un objet proche aura ses extrémités très éloignées d'elle. Cet indice monoculaire permet donc d'estimer visuellement la distance absolue à laquelle se trouve un objet et suppose d'avoir un champ de vision dégagé.
2. **Les dimensions relatives des objets** : chaque objet envoie au fond de l'œil une image dont les dimensions sont proportionnelles à celle de l'objet et qui décroissent avec une augmentation de la distance de l'objet par rapport à l'œil. Dans le cas où le cerveau connaît les dimensions "normales" de l'objet étudié, il peut alors en déduire sa distance. Dans le cas contraire, cet indice monoculaire ne permet que de réaliser des jugements relatifs de distance. Cet indice ne nécessitant qu'un *bas-niveau* de traitement est ainsi conceptuellement similaire à l'*intensité sonore* qui permet d'effectuer des jugements auditifs relatifs de distances.
3. **Effets de l'atmosphère** : plus les objets visuels sont éloignés, moins ils sont distincts à cause de l'humidité et des différents polluants présents dans l'atmosphère entre l'objet et l'observateur. Cet indice monoculaire permet alors une estimation absolue de la distance d'un objet pour des distances relativement importantes (> 100 m). Cet indice de *haut-niveau* est le pendant visuel de l'indice auditif nommé *absorption de l'air* présenté précédemment.
4. **Disparité binoculaire** : cet indice représente les différences de positions relatives d'un objet lorsque son image est projetée sur les rétines des deux yeux. L'analyse de cette disparité par le cerveau nous renseigne ainsi sur la distance absolue de l'objet observé. Cependant, étant donné que cet angle ne prend des valeurs notables que si les objets sont près des yeux alors cet indice binoculaire de *bas-niveau* est efficace surtout à des distances inférieures à quelques mètres.
5. **L'accommodation** : cet indice résulte de la déformation de la lentille des yeux permettant d'effectuer la mise au point (focus) sur des objets proches ou lointains tout en conservant nette l'image projetée sur la rétine. Cette information de *bas-niveau* peut alors être utilisée comme indice monoculaire de la perception de la distance absolue d'un objet visuel.

6. **La convergence** : il s'agit de l'angle entre les axes optiques de chacun des yeux. Les axes optiques se croisent à la distance où est situé l'objet qui est observé. Cette information de *bas-niveau* constitue un indice binoculaire de la perception de la distance absolue d'un objet.
7. **La parallaxe due au mouvement** : quand un observateur se déplace, les images rétinienne d'objets immobiles ont des mouvements et des variations de taille relatives par rapport à l'utilisateur dépendant des distances où se trouvent ces objets. L'amplitude des mouvements relatifs de ces objets permet alors à l'observateur d'apprécier la distance absolue à laquelle il se situe par rapport aux objets. Cet indice de *haut-niveau*, qui a aussi une contrepartie auditive, est un indice monoculaire de perception de la distance absolue d'un objet visuel.

Sur la figure 7.8 sont représentés pour chacun des indices mentionnés précédemment les seuils de détectabilité pour l'estimation de la distance entre deux objets situés respectivement à des distances D_1 et D_2 , normalisés par la distance moyenne à laquelle se trouvent ces objets, en fonction de cette distance moyenne. Cette figure permet d'avoir un aperçu synthétique de l'importance relative de chaque indice en fonction de la distance concernée : plus le seuil de détectabilité est bas, plus l'indice est sensible.

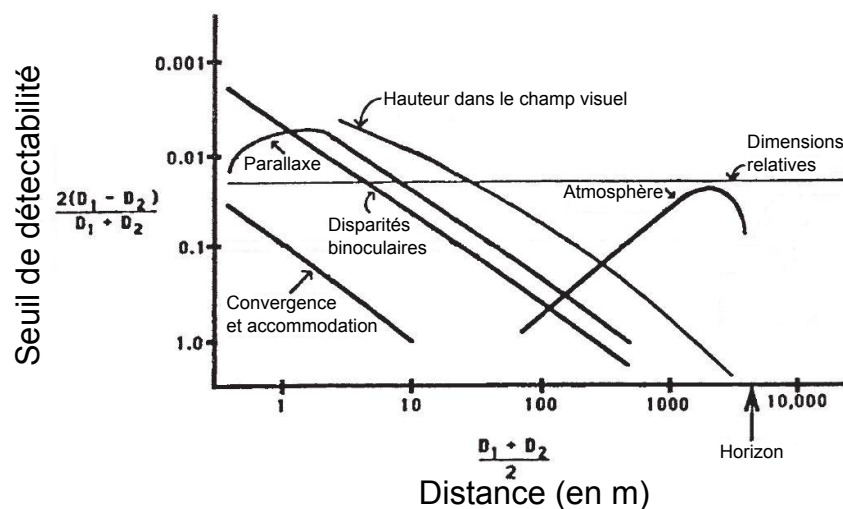


FIG. 7.8 – Ordonnée : Seuils de détectabilité pour chacun des indices mentionnés précédemment pour l'estimation de la distance entre deux objets situés respectivement à des distances D_1 et D_2 , normalisés par la distance moyenne à laquelle se trouvent ces objets. Abscisse : Distance moyenne à laquelle se trouvent ces objets. Figure réalisée d'après [Cutting 1997].

7.3 Intégration spatio-temporelle audio-visuelle

Dans cette section, la façon dont les flux provenant des modalités auditive et visuelle fusionnent de façon à donner lieu à la perception d'un objet audio-visuel localisable dans l'espace et dans le temps plutôt qu'à la perception d'un objet auditif et d'un objet visuel spatialement ou temporellement dissociés est analysée [Kohlrausch & van de Par 2005, Spence 2007].

7.3.1 Fenêtre spatio-temporelle d'intégration audio-visuelle

Même si la vision et l'audition travaillent d'un commun accord et ne sont pas indépendantes l'une de l'autre, elles sont pourtant différentes sur un certain nombre de points :

- **Aspects temporels** : lorsqu'un objet visible émet un son, les informations visuelles et sonores ont la même origine physique (l'objet), mais ne se propagent pas de la même façon. En effet, alors que la vitesse de la lumière dans l'air est d'environ $3 \times 10^8 \text{ ms}^{-1}$, la vitesse de propagation du son dans l'air n'est que de 340 ms^{-1} . De ce fait, les informations audio et visuelles n'arrivent pas simultanément aux niveaux des yeux et des oreilles. D'autre part, le traitement auditif est un traitement mécanique bas niveau réalisé relativement vite (i.e. $\simeq 1 \text{ ms}$), alors que le traitement visuel mettant en jeu des déclencheurs photochimiques est plus long (de l'ordre de 50 ms, soit $\simeq 17 \text{ m}$ parcourus par une onde sonore dans l'air) [Kopinska & Harris 2004, Burr & Alais 2006]. Les informations provenant des modalités visuelles et auditives arrivent donc avec des échelles temporelles différentes au cerveau qui les traite par conséquent inévitablement selon des échelles temporelles différentes.
- **Aspects spatiaux** : alors que la vue est très précise dans une petite zone de l'espace sur laquelle le regard est fixé, l'ouïe permet au contraire la localisation d'objets dans tout l'espace, mais de façon moins précise [Godfroy et al. 2011] (voir aussi figure 7.9). Les espaces auditifs et visuels sont donc représentés de façons différentes dans le cerveau. D'autre part, le système visuel étant totalement tributaire de la position des yeux, il est naturellement oculo-centré. Le système auditif base quant à lui sa représentation de l'espace sur les deux capteurs statiques dont il dispose (les oreilles) et est donc plutôt auriculo-centré. Le cerveau crée donc un référentiel multimodal unique à partir de la position relative des référentiels auditifs et visuels pour faire cohabiter les espaces auditifs et visuels [Hartnagel 2007].

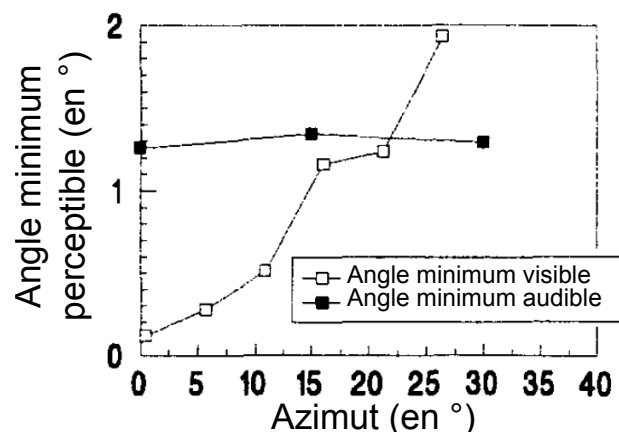


FIG. 7.9 – Angles minimums audible et visible en fonction de l'azimut, d'après Perrott [1993].

- **Codage de l'information** : une dernière différence entre l'ouïe et la vue réside dans la façon dont les informations relatives à l'espace sont codées [Fuchs et al. 2006a]. Le codage de l'information spatiale visuelle est directement réalisé par les bâtonnets et les cônes répartis sur l'ensemble de la rétine. Dans le cas de la vision, l'information spatiale est donc codée au plus bas niveau. Le codage de l'information sonore est réalisé par les cellules ciliées présentes dans la cochlée. En fonction de la fréquence, différentes zones de la cochlée sont excitées et donc différentes cellules ciliées sont

excitées. Le codage de l'information sonore est donc par essence temporel. Ce n'est ensuite qu'à un plus haut-niveau que les informations relatives à l'espace sonore sont déduites des informations temporelles (voir section 7.2.1).

Ces différences obligent ainsi le cerveau humain à artificiellement mettre en relation les flux d'informations provenant des modalités auditive et visuelle. Le cerveau doit alors faire preuve d'une certaine plasticité et tolérer des écarts spatio-temporels entre les flux auditif et visuel car il ne sait pas si l'origine des ces écarts est liée aux stimuli physiques qu'il reçoit ou bien à la captation et à l'interprétation qu'il fait de ces stimuli. Pour résumer, on peut citer Hartnagel [2007] qui estime que *“la perception spatiale n'est pas ancrée dans la spatialité physique de l'information [...] mais issue d'un réajustement dynamique dépendant [...] des systèmes de codage”*. Cela plaide en faveur de l'existence d'une fenêtre spatio-temporelle d'intégration audio-visuelle au sein de laquelle des écarts spatiaux et/ou temporels entre les stimuli auditif et visuel ne sont pas interprétables, et donc pas perçus.

Intégration temporelle

La plasticité du cerveau humain lors de l'intégration temporelle d'un stimulus audio et d'un stimulus visuel a été étudiée de façon intensive (voir par exemple [Hollier et al. 1999, Lewald et al. 2001, Lewald & Guski 2003]). Pour des stimuli audio-visuels simples (bruits blancs ou sinusoïdes pures accompagnés de flashes visuels), les fenêtres pour lesquelles les stimuli sont perçus comme étant temporellement alignés varient entre $[-25 \text{ ms}, +25 \text{ ms}]$ et $[-60 \text{ ms}, +160 \text{ ms}]$ (avec pour origine du temps l'instant de présentation du stimulus visuel) selon les stimuli et les protocoles employés. Pour des stimuli ayant un sens bien plus prononcé (comme une voix et un visage) cette fenêtre s'avère être plus large encore $[-90 \text{ ms}, +180 \text{ ms}]$. La fenêtre d'intégration temporelle audio-visuelle n'est donc pas symétrique et l'intégration se fait moins bien lorsque le stimulus auditif est en avance sur le stimulus visuel que lorsque c'est l'inverse. L'intégration audio-visuelle temporelle se fait par ailleurs de façon relativement indépendante de l'intégration en azimut [Wallace et al. 2004, Vroomen & Keetels 2006]. La figure 7.10 présente un récapitulatif du phénomène d'intégration audio-visuelle temporelle.

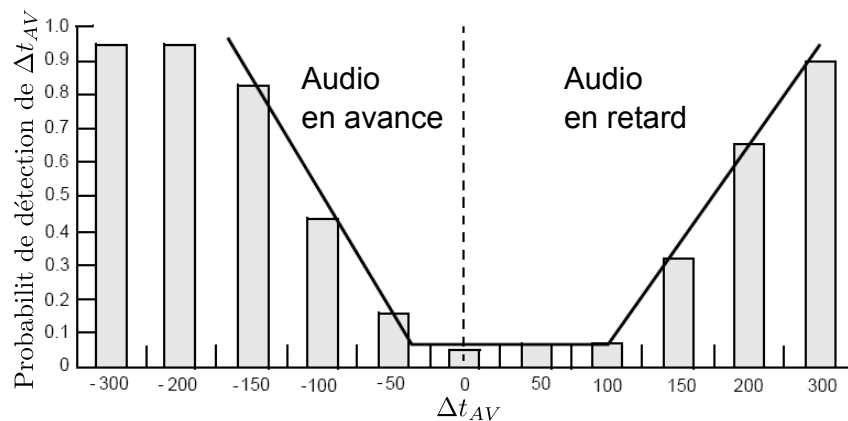


FIG. 7.10 – Probabilité de détection d'une erreur de synchronisation entre les stimuli auditifs et visuels en fonction du décalage entre eux, d'après Hollier et al. [1999].

Intégration en azimut

L'azimut est une autre dimension assez aisée à étudier dans le cadre de l'intégration audio-visuelle. Les études psycho-physiques consacrées à ce sujet et basées sur des stimuli abstraits situés sur l'axe médian (voir figure 7.1) révèlent des fenêtres spatiales d'intégration audio-visuelle en azimut dans le plan horizontal de l'ordre de 2 à 3° [Perrott 1993, Lewald et al. 2001, Lewald & Guski 2003, Godfroy et al. 2003, Hairston et al. 2003, Wallace et al. 2004]. Il est intéressant de noter que cette valeur est proche de la capacité de localisation humaine dans le domaine auditif dans le plan horizontal (voir Sec. 7.2.1). D'autres études, plutôt orientées vers des applications technologiques audio-visuelles (cinéma, télévision, etc.) se sont aussi intéressées à cette problématique [Komiya 1989, Sporer & Klehs 2004, Soto et al. 2008]. Ces études montrent que des écarts allant de 11° jusqu'à 30° sont tolérables pour des stimuli associant un visage animé et une voix. La fenêtre angulaire conduisant à une intégration audio-visuelle en azimut est donc assez dépendante des stimuli utilisés. Il est important de souligner ici que pour des stimuli abstraits, cette fenêtre, relativement étroite ($\simeq 3^\circ$) s'élargit lorsque les stimuli s'écartent de l'axe médian ($\simeq 5^\circ$) [Godfroy et al. 2003].

Intégration en élévation

L'intégration audio-visuelle en élévation est un sujet moins traité par la littérature. Godfroy et al. [2003] mettent en évidence l'influence de la direction de provenance des stimuli auditif et visuel sur l'étendue de la *zone d'intégration audio-visuelle*. À la différence de la *fenêtre d'intégration audio-visuelle* qui définit la fraction de l'espace au sein de laquelle une incohérence audio-visuelle n'est pas perçue, la zone d'intégration audio-visuelle obtenue ici correspond à l'écart en élévation ou en azimut à partir duquel 50% des personnes estiment que les stimuli audio et visuel sont incohérents spatialement. Pour le plan médian, des zones d'intégration de $13.1^\circ \pm 6.9^\circ$ en azimut et de $21.6^\circ \pm 10.1^\circ$ en élévation sont obtenues (stimulus auditif : salve de 500 ms de bruit rose, stimulus visuel : spot lumineux). Dans le plan médian, l'intégration audio-visuelle se fait donc plus facilement en élévation qu'en azimut, ce qui semble cohérent avec le fait que l'acuité de localisation auditive soit moins bonne dans la direction verticale que dans la direction horizontale. En dehors du plan médian l'anisotropie précédente disparaît et les zones d'intégration sont de dimensions similaires dans les deux directions. Des zones d'intégration audio-visuelles de $27^\circ \pm 9^\circ$ sont obtenues pour des stimuli audio placés à $\pm 20^\circ$ en azimut et en élévation.

Ces zones ne prédisent pas quel écart en azimut et en élévation entre les stimuli auditif et visuel conduit à une perte de l'intégration audio-visuelle (fenêtre d'intégration audio-visuelle). Cependant dans le plan médian, la zone d'intégration audio-visuelle semble être plus grande d'un facteur 1.5 environ en élévation qu'en azimut [Godfroy et al. 2003]. En se basant sur la valeur obtenue par Lewald et al. [2001] pour l'intégration en azimut et sur le ratio de 1.5 mis en évidence par Godfroy et al. [2003], des stimuli auditif et visuel abstraits seraient alors intégrés en élévation à la condition qu'ils ne soient pas espacés de plus de 5° . Cette extrapolation est cependant purement spéculative et doit être validée d'un point de vue perceptif.

Intégration en distance

Dans ce qui précède, nous avons évoqué les fenêtres spatiales d'intégration audio-visuelle permettant de positionner un objet par rapport au sujet selon deux des trois dimensions de l'espace. La troisième

dimension d'importance est la distance r à laquelle se situe l'objet par rapport à l'observateur.

Pour des stimuli très brefs, la différence de vitesses de propagation entre son et lumière semble être l'un des points clé de l'intégration audio-visuelle en distance. Différentes études [Sugita & Suzuki 2003, Kopinska & Harris 2004, Burr & Alais 2006] ont mis en évidence une relation linéaire existant entre la distance "source/sujet" et le décalage temporel correspondant à la simultanéité audio-visuelle perçue par les sujets. La pente de ces droites est de l'ordre de 1 à 3 ms/m. Cette valeur correspond donc relativement bien à $1/c = 2.924 \text{ ms/m}$ où $c = 342 \text{ m/s}$ est la vitesse de propagation du son dans l'air. Ainsi, une hypothèse expliquant l'intégration audio-visuelle en distance pour des stimuli brefs est que le cerveau humain serait capable d'estimer le décalage *plausible* (compte tenu de la distance de l'objet audio-visuel) entre la composante auditive et la composante visuelle, et de le compenser avant de décider de la simultanéité des stimuli. La fenêtre d'intégration temporelle serait donc simplement décalée en fonction de la distance dont proviennent les stimuli. Cette hypothèse reste cependant encore controversée (voir [Lewald & Guski 2004, Heron et al. 2007] par exemple).

Pour des stimuli plus longs, la fenêtre d'intégration en distance va dépendre de l'efficacité avec laquelle le cerveau va être capable de combiner les indices de localisation audio-visuels lui permettant d'estimer la distance des stimuli. Dans le cas où très peu d'indices de localisation auditifs sont disponibles pour estimer la distance (chambre anéchoïque), il a été montré que la fenêtre d'intégration en distance existait et que les stimuli audio-visuels étaient perçus à la position du stimulus visuel [Gardner 1968]. Ces résultats ont par la suite été nuancés lors d'études concernant des cas plus réalistes, incluant entre autres de la réverbération [Mershon et al. 1980, Zahorik 2001, Agganis et al. 2010, Bowen 2010]. Ainsi, même si aucune étude n'a véritablement cherché à établir les limites de la fenêtre d'intégration audio-visuelle en distance pour des stimuli longs, les études mentionnées précédemment laissent penser que cette fenêtre est de l'ordre de 50 cm à des distances de quelques mètres.

7.3.2 Facteurs cognitifs influents

En plus des facteurs spatio-temporels évoqués à la section 7.3.1, différents facteurs cognitifs contribuent à la facilité avec laquelle l'intégration audio-visuelle est réalisée [Spence 2007] :

1. **Le lien sémantique existant entre les stimuli auditifs et visuels** : la relation sémantique existant éventuellement entre les stimuli auditifs et visuels participe à la facilité avec laquelle le cerveau humain va intégrer les deux flux qui lui sont présentés. Par exemple, le lien sémantique existant entre l'image animée d'un visage et une voix est bien plus fort que celui existant entre une impulsion de bruit blanc et un spot lumineux. Pour la fenêtre d'intégration en azimuth, cette différence d'interprétation sémantique se traduit par une intégration audio-visuelle beaucoup plus tolérante pour le visage animé et la voix ($\simeq 30^\circ$) que pour les stimuli abstraits ($\simeq 3^\circ$).
2. **L'hypothèse d'unité** : le fait que le sujet s'attende à percevoir des stimuli issus d'une cause commune ou non influence la facilité avec laquelle l'intégration audio-visuelle est réalisée. Par exemple, une consigne du type "*Les objets visuels et sonores sont-ils cohérents entre eux ?*" laisse sous-entendre qu'il existe une relation entre les deux objets et donc va implicitement pousser les sujets à l'intégration audio-visuelle. À l'opposé, une consigne du type "*Y'a-t-il une différence spatiale entre les*

stimuli audio et visuels ?” suggère que les deux objets sont spatialement séparés, et donc va nuire à l’intégration audio-visuelle.

3. **La répartition des ressources attentionnelles** : les ressources attentionnelles dont dispose le cerveau humain sont à répartir, entre autres, entre les modalités auditive et visuelle. Il est cependant possible de basculer toute ou presque l’attention disponible sur l’une ou l’autre des modalités, au détriment de l’autre. Le fait d’orienter les sujets vers une tâche de localisation visuelle ou auditive, ou de ne rien préciser et donc de laisser le sujet faire le choix, est susceptible de provoquer un basculement de l’attention vers l’une ou l’autre des modalités au détriment de l’intégration audio-visuelle.

Les facteurs cognitifs influençant l’intégration audio-visuelle sont donc nombreux et difficilement contrôlables. La généralisation des résultats précédents est donc à nuancer.

7.3.3 Conséquences

Différentes illusions audio-visuelles résultent de l’ambiguïté perceptive créée par l’intégration audio-visuelle [Kohler & van de Par 2005, Spence 2007]. On peut citer par exemple :

1. **L’interprétation d’une scène audio-visuelle ambiguë** : la façon dont sont intégrés dans le temps et dans l’espace les différents stimuli composant une scène audio-visuelle va influencer le sens qui lui sera donné. Par exemple, lorsque deux sphères visuelles identiques sont sur la même trajectoire mais évoluent dans des directions opposées, si un son d’impact est émis lorsque les sphères se croisent et à la position où les sphères se croisent alors les sphères seront perçues comme rebondissant l’une contre l’autre. Si des sons continus sont attachés à chacune des sphères, elles seront par contre perçues comme deux objets indépendants continuant chacun sa route. Un autre exemple consiste à associer une série de bips sonores à un unique flash visuel. Dans ce cas, plusieurs flashes visuels sont généralement perçus [Shams et al. 2002].
2. **Effet ventriloque ou biais intermodal** : lorsque deux stimuli visuel et auditif sont émis avec des décalages temporels et spatiaux permettant l’intégration audio-visuelle, l’objet audio-visuel résultant de la fusion des deux stimuli est perçu à un instant proche de celui où le stimulus auditif est émis et à une position située à proximité de celle d’où le stimulus visuel provient [Jack & Thurlow 1973, Radeau 1994, Hairston et al. 2003, Wallace et al. 2004]. Dans les zones de l’espace où la modalité visuelle est très précise (le long de l’axe médian par exemple) la position de l’objet audio-visuel perçu est indifférentiable de celle de l’objet visuel, et ce même pour des stimuli audio relativement éloignés. On parle alors de *capture visuelle*. Cet effet est utilisé par les *ventriloques* lorsqu’ils veulent donner l’impression que leurs marionnettes parlent, ou bien par le cinéma ou la télévision pour donner l’impression que les voix des acteurs viennent des images projetées.
3. **Utilisation optimale des informations disponibles** : l’analyse constante des flux auditif et visuel redondants permet au cerveau humain d’extraire des informations plus précises en contexte bimodal qu’en contexte unimodal. Il semble ainsi que le cerveau utilise les données issues de chacun des sens et les pondère selon leurs fiabilités respectives selon un modèle de maximum de vraisemblance [Perrott 1993, Battaglia et al. 2003, Alais & Burr 2004, Nguyen 2008, Godfroy et al. 2011]. Dans le cas de la localisation spatiale, l’être humain est capable, dans des conditions expérimentales données,

de localiser un stimulus visuel situé à une position m_V avec une variance σ_V , et un stimulus audio situé à une position m_A avec une variance σ_A . Dans un contexte audio-visuel, un stimulus audio-visuel est alors localisé autour d'une position m_{AV} et avec une variance σ_{AV} données par [Alais & Burr 2004] :

$$m_{AV} = \frac{\sigma_V^2}{\sigma_V^2 + \sigma_A^2} m_A + \frac{\sigma_A^2}{\sigma_V^2 + \sigma_A^2} m_V \quad \text{et} \quad \sigma_{AV}^2 = \frac{\sigma_A^2 \sigma_V^2}{\sigma_A^2 + \sigma_V^2} \leq \min(\sigma_A, \sigma_V)$$

La variance σ_{AV} avec laquelle est localisé le stimulus audio-visuel est donc inférieure aux variances des deux modalités prises séparément. La position m_{AV} à laquelle est perçu l'objet audio-visuel est une pondération des estimations données par chacune des modalités en fonction de leur fiabilité. Dans le cas où la modalité visuelle est beaucoup plus fiable que la modalité auditive (*i.e.* quand $\sigma_V \gg \sigma_A$), alors $m_{AV} \simeq m_V$ et $\sigma_{AV} \simeq \sigma_V$: il s'agit du phénomène de capture visuelle décrit précédemment. En périphérie du champ de vision, quand $\sigma_A \simeq \sigma_V$ (voir figure 7.9), l'audition et la vision conduisent à des localisations de même variabilité et sont donc utilisées conjointement.

7.4 Évaluation des performances offertes par le SMART-I²

Une évaluation du SMART-I² a été réalisée dans le but d'estimer les limites imposées par le système sur la précision de localisation en azimuth, d'étudier la capacité de restitution de l'effet de parallaxe auditive et d'évaluer la cohérence perceptive du rendu audio-visuel. Cette évaluation a été réalisée sur une version préliminaire du dispositif composée d'un seul LaMAP (voir figure 7.11). Elle a par ailleurs fait l'objet d'une présentation lors de la 125^{ième} convention de l'*Audio Engineering Society* [Rébillat et al. 2008].

7.4.1 Présentation générale

Cette évaluation s'est attachée à mettre en relation des résultats subjectifs obtenus par le biais d'expériences perceptives et des résultats objectifs obtenus à partir de mesures acoustiques dans la zone de rendu. Ainsi, deux expériences perceptives et une campagne de mesures objectives ont été réalisées :

1. **Une expérience classique de localisation en azimuth** : des sujets statiques doivent estimer de quel haut-parleurs, parmi les haut-parleurs visuels virtuels positionnés sur un arc de cercle, provient le stimulus auditif qu'ils viennent d'entendre.
2. **Une expérience visant à valider le rendu de l'effet de parallaxe auditive** : des sujets sont face à une ligne de haut-parleurs visuels virtuels et doivent déterminer lequel d'entre eux est la source sonore en explorant la zone autorisée et en se basant uniquement sur l'effet de parallaxe auditive.
3. **Une campagne de mesures objectives** : en parallèle de ces expériences, le champ acoustique rayonné par chacun des 12 excitateurs de cette première version du dispositif a été mesuré à l'aide d'une tête artificielle (Neumann KU-100) dans le but d'évaluer objectivement la qualité des indices de localisation auditifs restitués par le système de rendu audio spatialisé dans la zone de rendu.

La figure 7.11 résume l'organisation spatiale du dispositif et les positions des excitateurs, des objets audio-visuels virtuels utilisés dans les expériences relatives à l'azimut et à la parallaxe, ainsi que les positions auxquelles a été mesuré le champ sonore à l'aide de la tête artificielle.

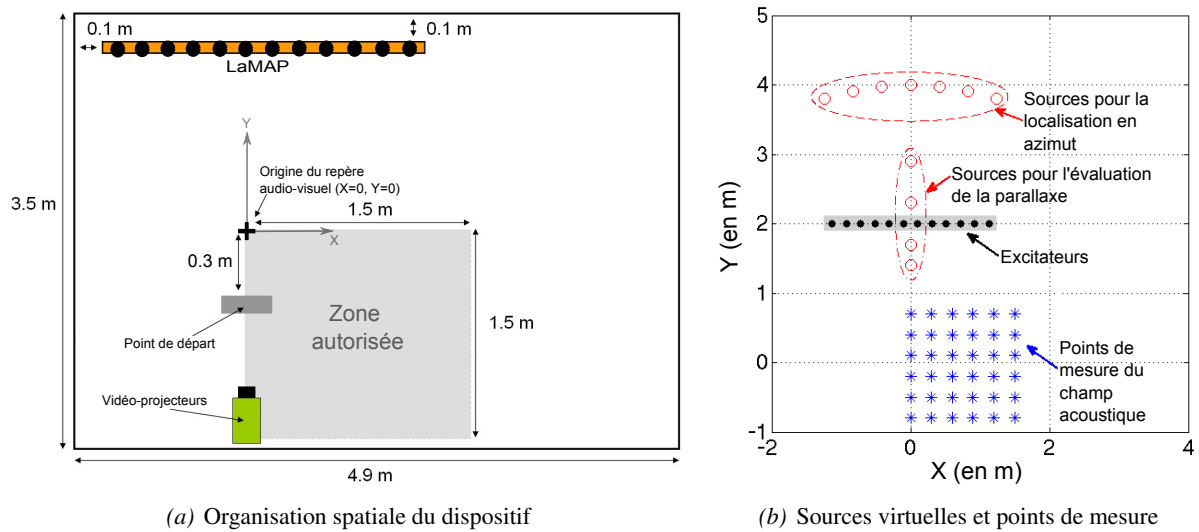


FIG. 7.11 – (a) : organisation spatiale du dispositif préliminaire. (b) : positions des excitateurs, des objets sonores virtuels utilisés dans les expériences relatives à l’azimut et à la parallaxe, ainsi que les positions auxquelles a été caractérisé le champ sonore à l’aide de la tête artificielle.

Les expériences perceptives se sont déroulées en trois phases pour une durée totale de 40 minutes par session :

1. Scène d’introduction ($\simeq 5$ min)
2. Localisation en azimuth ($\simeq 15$ min)
3. Restitution de l’effet de parallaxe ($\simeq 15$ min)

Quatorze sujets (9 hommes) âgés de 22 à 53 ans ont participé aux expériences. Tous ont affirmé avoir une vue normale (éventuellement corrigée) ainsi qu’une audition normale. Tous les sujets ont commencé leur session avec la scène d’introduction. Pour la moitié d’entre eux ils ont poursuivi avec la localisation en azimuth puis avec la restitution de l’effet de parallaxe. Pour l’autre moitié, l’ordre des deux expériences était inversé.

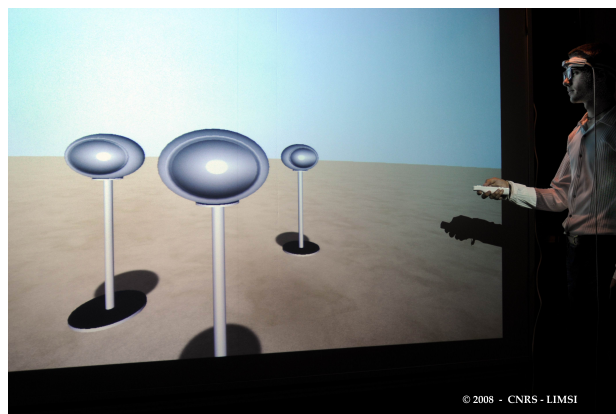


FIG. 7.12 – Un sujet lors de la scène d’introduction (Photo : P. Durand).

La scène d'introduction est composée de trois haut-parleurs visuels virtuels reposant sur un sol sableux (voir la figure 7.12). Seul l'un d'entre eux est activé et diffuse de la musique. Le sujet peut alors, à l'aide d'une Wiimote, sélectionner le haut-parleur visuel virtuel actif et se déplacer dans la zone de rendu pour se familiariser avec le rendu stéréoscopique et la WFS.

7.4.2 Analyse objective

L'analyse objective utilisée ici cherche à estimer les erreurs commises sur les indices de localisation auditifs (ITD et ILD) par le SMART-I² en se basant sur des réponses impulsionnelles binaurales mesurées et sur un modèle simple d'audition [Sanson et al. 2008]. Cet outil assez sommaire ne permet cependant en aucun cas de prédire la position perçue des sources virtuelles.

Une tête artificielle (Neumann KU-100) a été positionnée à 36 positions régulièrement espacées (30 cm) selon X et Y dans la zone de rendu (voir figure 7.11). Au cours de ces mesures, la tête artificielle est dirigée dans la direction Y et fait face au LaMAP.

Les erreurs commises sur les indices de localisation auditifs (ITD et ILD) sont évaluées en 5 étapes⁴ récapitulées sur la figure 7.13 [Sanson et al. 2008] :

1. Mesures en chambre anéchoïque des HRTFs de la tête artificielle utilisée.
2. Estimation des réponses impulsionnelles binaurales “optimales” correspondant à chacune des sources virtuelles. Ces réponses impulsionnelles sont obtenues en utilisant les HRTFs de la tête artificielle : ce sont les sources virtuelles “optimales”.
3. Mesures sur site des réponses impulsionnelles binaurales des 12 excitateurs à différentes positions dans la zone de rendu.
4. Estimation des réponses impulsionnelles binaurales “réalistes” correspondant à chacune des sources virtuelles. Ces réponses impulsionnelles sont obtenues en sommant les contributions (mesurées précédemment) des 12 excitateurs pilotés par WFS (voir section 2.5.3) : ce sont les sources virtuelles “réalistes”.
5. Détermination des indices de localisation auditifs (ITD et ILD) correspondant aux sources virtuelles “optimales”, $ITD_{id}(f_n)$, $ILD_{id}(f_n)$, et “réalistes”, $ITD_{re}(f_n)$, $ILD_{re}(f_n)$, dans $N = 40$ bandes fréquentielles réparties entre 100 Hz et 20 kHz et centrées sur les fréquences $\{f_n\}_{n \in [1, N]}$ en utilisant un modèle simple d'audition⁵.
6. Calcul des erreurs $\Delta ITD(f_n) = ITD_{re}(f_n) - ITD_{id}(f_n)$, et $\Delta ILD(f_n) = ILD_{re}(f_n) - ILD_{id}(f_n)$, commises dans le cas “réaliste” en comparaison au cas “optimal” dans chacune des bandes de fréquence étudiées.

Des critères globaux indépendants de la fréquence peuvent ensuite être calculés à partir de $\{\Delta ITD(f_n)\}_{n \in [1, N]}$ et de $\{\Delta ILD(f_n)\}_{n \in [1, N]}$ en utilisant une fonction de pondération normalisée $q(f)$ rendant compte l'importance relative des différentes bandes fréquentielles dans les tâches de localisation [Stern & Shear 1996, Sanson et al. 2008].

4. Toute la phase de traitement du signal relative à cette partie du travail a été réalisée par É. Corteel.

5. L'ILD est calculé comme la différence d'énergie entre les signaux présentés aux deux oreilles. L'ITD est calculé en se basant sur le maximum de la fonction de corrélation inter-aurale. Les détails du modèle utilisé sont décrits dans [Sanson et al. 2008].

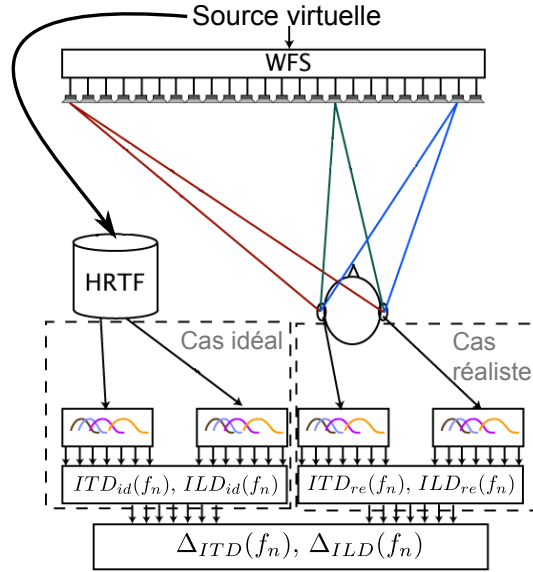


FIG. 7.13 – Calcul des erreurs commises sur les indices de localisation auditifs par le système réaliste en comparaison au système optimal.

Le biais d'ITD et la variabilité d'ITD sont alors définis comme :

$$B_{ITD} = \text{mean} [\Delta_{ITD}(f_n)q(f_n)]_{n \in [1, N]} \quad \text{et} \quad V_{ITD} = \text{std} [\Delta_{ITD}(f_n)q(f_n)]_{n \in [1, N]} \quad (7.1)$$

Le biais d'ILD et la variabilité d'ILD sont de plus définis comme :

$$B_{ILD} = \text{mean} [\Delta_{ILD}(f_n)q(f_n)]_{n \in [1, N]} \quad \text{et} \quad V_{ILD} = \text{std} [\Delta_{ILD}(f_n)q(f_n)]_{n \in [1, N]} \quad (7.2)$$

Les biais B_{ITD} et B_{ILD} indiquent l'écart moyen existant entre les indices de localisation optimaux et les indices de localisation effectivement reçus par le sujet. Ils peuvent être interprétés comme les indications d'une éventuelle dérive perceptive en azimuth causée par des indices de localisation incorrects créés par le système de reproduction. La convention qui est choisie ici est telle que les biais engendrés sont dirigés vers la droite pour des valeurs positives de B_{ITD} et B_{ILD} . Les variabilités V_{ITD} et V_{ILD} sont des indicateurs de la cohérence des indices de localisation au sein des différentes bandes fréquentielles et donc de la *localisabilité* de la source virtuelle. Il est important de noter que les critères globaux présentés ici font sens uniquement dans des situations où différents cas de figure sont comparés. Étant donnée la simplicité du modèle de localisation utilisé, ce ne sont en aucun cas des prédictions absolues de localisation.

Les biais B_{ITD} et B_{ILD} et les variabilités V_{ITD} et V_{ILD} ont été calculés à partir des mesures réalisées dans la zone d'écoute du SMART-I² pour toutes les sources virtuelles utilisées dans les expériences perceptives et pour tous les points de la zone d'écoute où des mesures ont été effectuées (voir figure 7.11). Ces critères globaux ont aussi été calculés, pour les mêmes combinaisons de sources virtuelles et de position d'écoute, pour un dispositif de WFS optimal composé de 12 sources omnidirectionnelles placées dans un environnement anéchoïque aux mêmes positions que les excitateurs du SMART-I². La comparaison de ces différents critères permet donc de quantifier les défauts de restitution sonore spatialisé engendrés par le SMART-I² en comparaison d'un dispositif optimal pour la WFS.

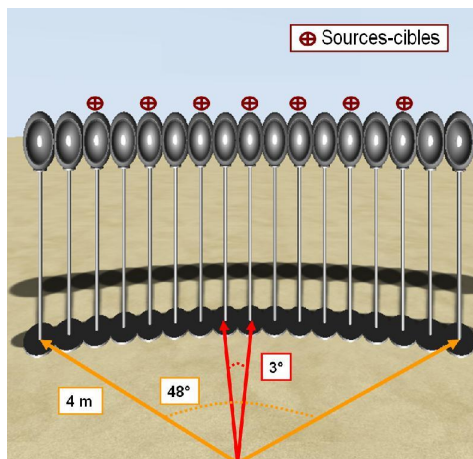
7.4.3 Précision de localisation en azimuth

L'objectif de cette expérience est d'évaluer la précision de localisation en azimuth qu'autorise le rendu sonore spatialisé créé par le SMART-I².

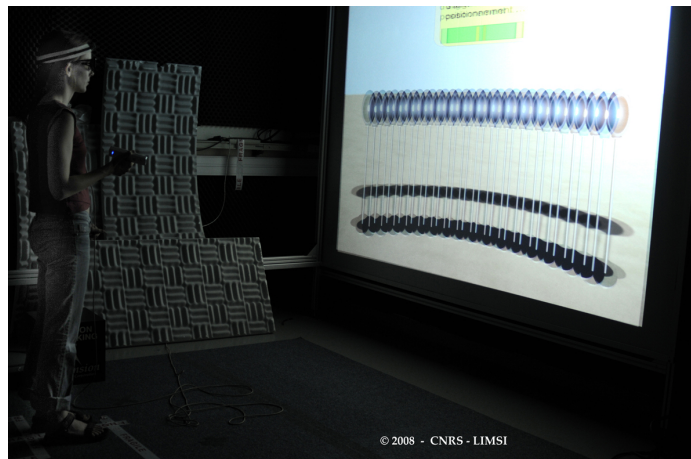
Protocole expérimental

La scène visuelle utilisée est constituée de 17 haut-parleurs virtuels disposés sur un arc de cercle de 4 m de rayon et centré sur le sujet (voir figure 7.14). Les haut-parleurs virtuels sont séparés les uns des autres par un azimuth de 3°. Cette valeur d'écartement a été choisie car elle est de l'ordre de grandeur des capacités de localisation humaines en azimuth (voir section 7.2.1). Il n'est donc pas nécessaire d'avoir davantage de haut-parleurs visuels. L'ouverture angulaire totale est de 48°.

Parmi les 17 haut-parleurs virtuels présents, seuls 7 peuvent être auditivement actifs, ce sont les sources-cibles. Les autres haut-parleurs servent de support visuel pour le report de l'azimuth perçu. Cette méthode de report a été choisie car elle permet de voir si les sujets sont capables de transporter correctement une information acquise dans l'espace auditif jusque dans l'espace visuel. Le but est de savoir si les composantes audio et visuelle du rendu audio-visuel proposé sont ou non perceptivement cohérentes entre elles et avec l'espace réel.



(a) Organisation spatiale de la scène



(b) Un sujet lors de l'expérience : phase de contrôle de sa position
(photo : P. Durand)

FIG. 7.14 – Scène audio-visuelle utilisée pour l'évaluation de la localisation en azimuth. (a) : les 17 haut-parleurs virtuels ainsi que les 7 potentielles cibles acoustiques et les dimensions spatiales de la scène sont indiquées (informations non-fournies aux sujets). (b) : un sujet est en train de passer l'expérience.

Le stimulus auditif utilisé est une impulsion de bruit blanc longue de 150 ms et possédant des durées de montée et de descente égales à 10 ms. Ce stimulus large bande permet aux sujets d'utiliser tous les indices disponibles pour la localisation en azimuth (voir section 7.2.1). Ce stimulus est joué une seule fois à un niveau de 68 dBA au point de départ. Le niveau de bruit de fond à la position du sujet est de 40 dBA. Au cours de l'expérience, le stimulus auditif est répété 15 fois par chacune des 7 sources-cibles, ce qui donne un total de 105 essais.

Au début de l'expérience, le sujet est positionné au point de départ (voir figure 7.11). Il ne lui est pas autorisé de se déplacer durant cette expérience. Chaque essai est organisé de la façon suivante :

1. Le sujet indique qu'il est prêt à effectuer la tâche de localisation en appuyant sur un bouton de la Wiimote.
2. La position du sujet et l'orientation de sa tête sont contrôlées par une procédure automatisée : le sujet doit être au point de départ et doit fixer la direction Y (voir figure 7.11).
3. Si le sujet est prêt et que sa position et son orientation sont correctes, le stimulus auditif est joué. Le stimulus auditif n'est joué qu'une seule fois.
4. Un curseur de sélection visuel apparaît au-dessus d'un des haut-parleurs virtuels, et le sujet indique de quel haut-parleur virtuel il pense que provient le stimulus auditif en déplaçant le curseur jusqu'au haut-parleur virtuel correspondant. Il valide son choix en appuyant à nouveau sur un bouton de la Wiimote.

Dans cette expérience, la variable indépendante est la position angulaire des sources-cibles. Les variables dépendantes sont l'azimut estimé par les sujets et le temps de localisation.

Résultats de l'évaluation subjective

Les résultats de l'évaluation subjective sont présentés sur la figure 7.15. Les angles sont donnés dans le sens horaire.

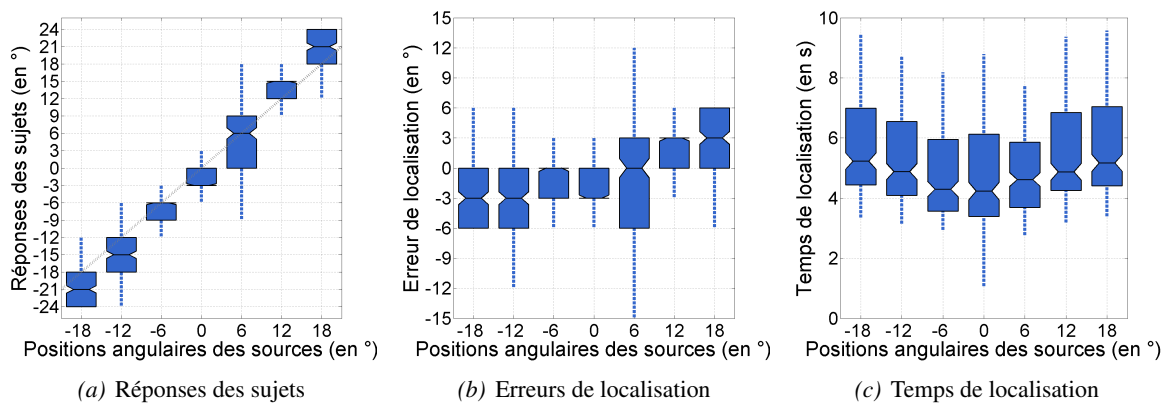


FIG. 7.15 – Résultats de l'expérience de localisation en azimut. Les angles sont donnés dans le sens horaire. Sur la figre 7.15(a), la ligne pointillée grise indique la droite d'équation $y = x$.

Les réponses des sujets ont été analysées à l'aide d'une analyse de variance à mesures répétées avec un niveau de significativité de 5% [Rietveld & van Hout 2005] (abrégé ANOVA par la suite) prenant les positions angulaires des sources comme facteur. Les résultats obtenus indiquent un effet très significatif de la position angulaire de la source sur les réponses des sujets ($F(6) = 112$ et $p < 10^{-5}$). Un test post-hoc (t -test avec correction de Bonferroni) révèle de plus que chacune des positions angulaires des sources est significativement différente des autres. Les différentes sources cibles ont donc toutes été correctement distinguées les unes de autres, comme le montre la figure 7.15(a).

Les erreurs de localisation faites par les sujets ont aussi été analysées à l'aide d'une ANOVA à mesures répétées prenant les positions angulaires des sources comme facteur. Les résultats obtenus n'indiquent aucun effet significatif de la position angulaire de la source sur les réponses des sujets ($F(6) = 1.8$ et $p = 0.11$). Les erreurs de localisation réalisées lors de la localisation de chacune des sources-cibles sont donc comparables, comme le montre la figure 7.15(b). L'erreur moyenne de localisation en azimuth est ici de $-0.9 \pm 5.1^\circ$. L'erreur moyenne absolue de localisation en azimuth est de $3.7^\circ \pm 3.5^\circ$. Une légère tendance à surestimer les azimuths des sources extrêmes est remarquable sur les figures 7.15(a) et 7.15(b), mais n'est cependant pas significative dans l'analyse.

Les durées de localisation des sujets ont enfin été analysées à l'aide d'une ANOVA à mesures répétées prenant les positions angulaires des sources comme facteur. Les résultats obtenus n'indiquent aucun effet significatif de la position angulaire de la source sur les réponses des sujets ($F(6) = 0.27$ et $p = 0.95$). Les temps de localisation mis pour la localisation de chacune des sources-cibles sont donc comparables, comme le montre la figure 7.15(c). Les sources-cibles sont localisées en moyenne en 5.8 ± 3.7 secondes. On constate une légère tendance des sujets à estimer plus rapidement les azimuths des sources situées près de 0° que ceux des sources latérales. Cela peut être attribué au fait que le curseur de sélection apparaissait toujours à 0° , et donc que les sujets avaient moins à le déplacer pour des sources-cibles situées autour de ce point que pour les sources-cibles latérales.

Résultats de l'évaluation objective

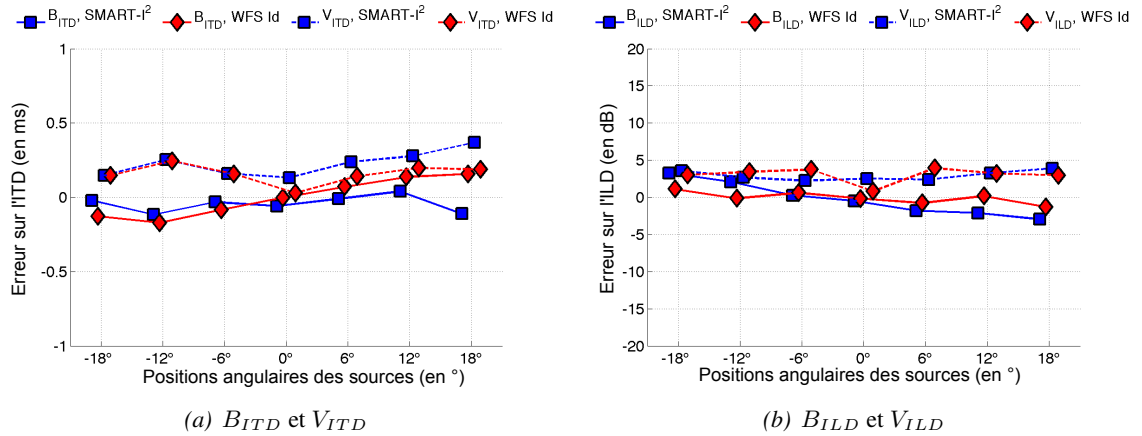


FIG. 7.16 – B_{ITD} et V_{ITD} (à gauche), B_{ILD} et V_{ILD} (à droite) déterminés à la position du sujet (point de départ) pour le SMART-I² et le dispositif optimal pour la WFS. Les échelles verticales choisies sont représentatives des ITD et ILD attendus par le système auditif humain.

Les biais et variabilités des ITD et ILD ont été déterminés à la position du sujet (point de départ, voir figure 7.11) pour le SMART-I² et pour le dispositif de WFS optimal, et sont présentés à la figure 7.16. Dans un premier temps, il est assez notable de constater que les erreurs réalisées sur les indices de localisation par la WFS au point de départ ($\simeq 0.1$ ms pour l'ITD et $\simeq 1$ dB pour l'ILD) sont relativement faibles. Pour avoir un élément de comparaison, une source située à 90° en azimuth crée un ITD $\simeq 1$ ms et un

ILD $\simeq 15$ dB, ce qui donne un ordre de grandeur des variations maximales d'ITD et d'ILD attendues par le système auditif [Blauert 1999]. D'autre part, les erreurs sur les indices de localisation induites par le SMART-I² sont assez similaires à celles réalisées par le dispositif de WFS optimal. Cependant, la variabilité de l'ITD est légèrement plus élevée pour le SMART-I² que pour le dispositif optimal. Enfin, le biais de l'ILD présente une courbe qui suggère une compression perceptive de l'azimut des sources virtuelles. Les résultats perceptifs présentés dans les figures 7.15(a) et 7.15(b) indiquent cependant plutôt une dilatation perceptive de l'azimut.

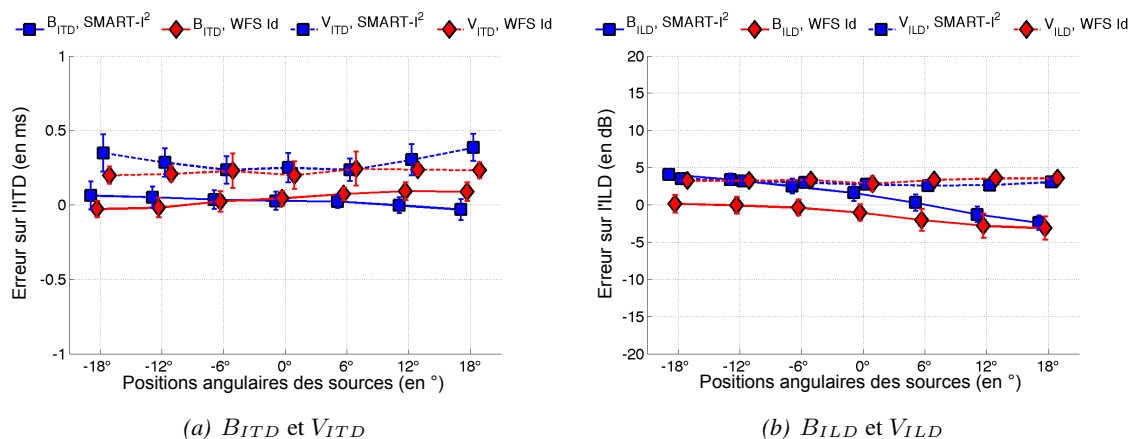


FIG. 7.17 – B_{ITD} et V_{ITD} (à gauche), B_{ILD} et V_{ILD} (à droite) moyennés sur toutes les positions de mesures pour le SMART-I² et le dispositif optimal pour la WFS. Les traits verticaux indiquent l'écart-type. Les échelles verticales choisies sont représentatives des ITD et ILD attendus par le système auditif humain.

La figure 7.17 présente les biais et variabilités de ITD et ILD moyennés sur toutes les positions de mesures pour le SMART-I² et le dispositif optimal pour la WFS. Les traits verticaux indiquent l'écart-type. Les biais moyens en ITD et ILD suivent des tendances similaires à celles observées pour la position centrale. De plus, les écart-types qui leur sont associés restent faibles, typiquement inférieurs à 0.1 ms pour le biais d'ITD et à 1 dB pour le biais d'ILD. La précision de localisation observée au point de départ est donc a priori de la même qualité dans toute la zone d'écoute. Cela confirme donc bien que la WFS n'est pas assujettie aux problèmes de *sweet spot* posés par d'autres technologies de rendu audio spatialisé (voir section 2.4.2).

Discussion

Le cerveau humain est capable de localiser des sources sonores en azimut avec une très bonne précision ($\simeq 0.1^\circ$) et une variabilité de allant de 1° à 4° en fonction du stimulus utilisé, comme on peut le voir par exemple sur la figure 7.4 [Blauert 1999, Godfroy et al. 2011]. Les résultats de l'étude présentée font apparaître une erreur moyenne signée de localisation en azimut de $-0.9^\circ \pm 5.1^\circ$. Les sujets sont donc capables de localiser une source sonore virtuelle assez précisément en azimut mais avec une variabilité qui est un peu plus importante que celle observée pour des sources sonores réelles. Cela peut être imputable à la WFS pour laquelle les variabilités V_{ITD} et V_{ILD} , qui donnent une idée de la cohérence des indices de

localisation au sein des différentes bandes fréquentielles et donc de la *localisabilité* de la source virtuelle, sont relativement élevées. La méthode de report utilisée ici, qui impose un choix discontinu par pas de 3° en azimut, est aussi susceptible d'augmenter artificiellement la variabilité des réponses des sujets.

Les résultats obtenus ici peuvent aussi être comparés avec des résultats obtenus par d'autres études utilisant la WFS. Verheijen [1998] a effectué une étude similaire à celle présentée ici, mais avec un banc de haut-parleurs électrodynamiques (et non des LaMAPs) et un report effectué sur des haut-parleurs réels (et non virtuels) espacés de $\simeq 2^\circ$. Il rapporte une erreur absolue de localisation en azimut de $3.5^\circ \pm 1.5^\circ$ [Verheijen 1998, Fig. 6.4]. Dans l'étude présentée ici, une erreur absolue de localisation en azimut de $3.7^\circ \pm 3.5^\circ$ est trouvée. Là encore, la valeur moyenne obtenue concorde avec les études précédentes, mais l'écart-type qui lui est associé est un peu plus important. Cette augmentation de la variabilité est peut-être due au rayonnement des LaMAPs qui est plus complexe que celui des haut-parleurs électrodynamiques classiques utilisés par Verheijen [1998]. Une autre explication concerne la méthode de report utilisée ici, qui impose un choix discontinu par pas de 3° au lieu de $\simeq 2^\circ$ pour Verheijen [1998].

7.4.4 Restitution de l'effet de parallaxe

L'objectif de cette expérience est d'évaluer la capacité du SMART-I² à restituer l'effet de parallaxe auditive.

Protocole expérimental

La scène audio-visuelle utilisée ici est composée de 8 haut-parleurs visuels virtuels disposés tous les 30 cm sur une ligne perpendiculaire au dispositif (voir figure 7.11). Pour les mêmes raisons que celles décrites dans l'expérience précédente, seuls 4 haut-parleurs parmi les 8 sont actifs auditivement : ce sont les sources-cibles (1.4 m, 1.7 m, 2.3 m and 2.9 m avec pour référence l'origine du repère $X = 0, Y = 0$). Étant donné que le premier haut-parleur de la ligne est situé 1.1 m devant l'origine du repère audio-visuel (voir figure 7.18), les 3 premiers haut-parleurs virtuels sont donc situés devant l'écran de projection et les sources-cibles situées à 1.4 m et à 1.7 m sont des sources focalisées.

Le stimulus auditif utilisé ici est une salve de bruit blanc d'une seconde, filtrée par un filtre passe-bas de fréquence de coupure égale à 4 kHz, modulé sinusoïdalement en amplitude à 18 Hz, et suivie par un silence d'une seconde. Ce stimulus large bande permet aux sujets d'utiliser un maximum d'indices de localisation relatifs à l'azimut lors de leurs déplacements dans la zone de rendu⁶. Contrairement à l'expérience précédente, le stimulus est répété jusqu'à ce que le sujet ait validé son choix. Tous les stimuli auditifs sont présentés à un niveau de 70 dBA au point de départ. Ainsi, l'intensité sonore perçue par le sujet

6. Il faut noter que le stimulus auditif utilisé ici contient de l'énergie au dessus de la fréquence temporelle de repliement spatial $f_{al} \simeq 1.2$ kHz du SMART-I², qui est la limite fréquentielle à partir de laquelle le champ sonore n'est ni spatialement, ni temporellement correctement reconstruit. Le filtrage passe-bas avec une fréquence de coupure située à 4 kHz a donc été réalisé dans le but de limiter la part d'énergie présente au dessus de f_{al} . Néanmoins il a été montré, pour des signaux à large bande spectrale, que l'ITD en basses fréquences joue un rôle dominant dans la localisation en azimut [Wightman & Kistler 1992]. De plus, Sanson et al. [2008] font la preuve expérimentalement que les champs sonores reconstruits par la WFS autorisent une localisation précise en azimut, même si de l'énergie est présente à des fréquences supérieures à la fréquence f_{al} . Et Cortel et al. [2007] montrent que, même s'ils ne sont pas exacts, les indices de localisation obtenus au dessus de f_{al} en utilisant des MAPs sont généralement cohérents avec les indices de localisation en basse fréquence restitués par la WFS. Pour ces raisons, il semble donc peu probable que la part d'énergie des signaux située au dessus de f_{al} soit responsable d'erreurs de localisation. Cela semble de plus confirmé par les résultats de l'analyse objective.

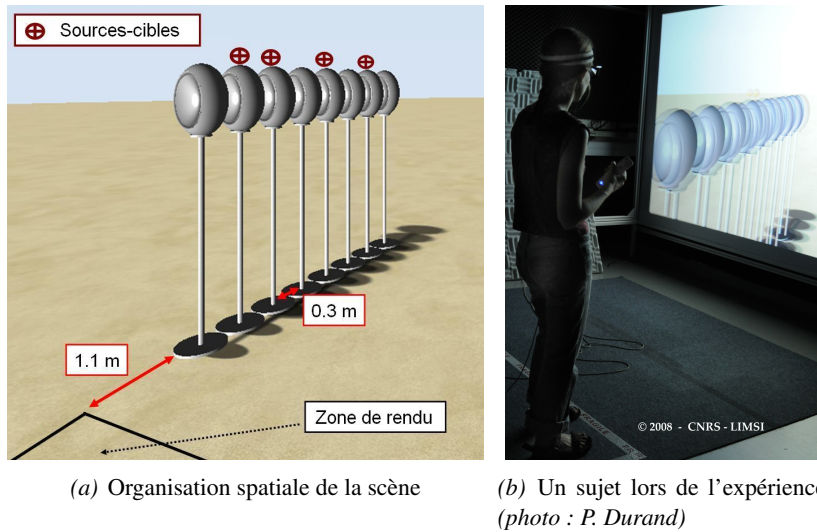


FIG. 7.18 – Scène audio-visuelle utilisée pour l'évaluation de restitution de l'effet de parallaxe. (a) : les 8 haut-parleurs virtuels ainsi que les 4 potentielles cibles acoustiques et les dimensions spatiales de la scène sont indiquées. (b) : un sujet est en train de passer l'expérience.

au début de chaque essai n'est pas informative quand à la distance à laquelle se trouve la source sonore. D'autre part, aucun effet de salle n'est ajouté. L'effet de parallaxe auditive est alors l'indice perceptif majoritairement utilisable ici pour l'estimation des distances (voir figure 7.3). Le stimulus auditif est joué 10 fois pour chacune des positions des sources-cibles, ce qui fait un total de 40 essais par session.

Durant cette expérience, les sujets sont encouragés à se déplacer dans la zone de rendu de façon à acquérir le plus possible d'informations concernant la source sonore active. L'utilisateur démarre chaque essai au point de départ (voir figure 7.11) et sa position est contrôlée avant que le stimulus ne commence à être joué.

Dans cette expérience, la position en mètres des sources-cibles est la variable indépendante. Les variables dépendantes sont la distance estimée par les sujets et le temps d'estimation.

Résultats de l'évaluation subjective

Les résultats de l'évaluation subjective correspondant à l'expérience étudiant la restitution de l'effet de parallaxe sont présentés à la figure 7.19.

Les réponses des sujets ont été analysées à l'aide d'une ANOVA à mesures répétées prenant les positions des sources comme facteur. Les résultats obtenus indiquent un effet très significatif de la position de la source sur les réponses des sujets ($F(3) = 7$ et $p < 10^{-3}$). Un test post-hoc (t -test avec correction de Bonferroni) révèle de plus que 3 groupes de sources sont significativement différenciés par les sujets : un groupe constitué de la source située à 1.4 m, un groupe constitué de celle située à 1.7 m, et un dernier constitué de celles situées à 2.3 m et 2.9 m. Les différentes sources-cibles ne sont donc pas forcément distinguées les unes de autres, en particulier les plus lointaines d'entre elles, comme le montre la figure 7.19(a). Les sujets perçoivent cependant quand même une certaine impression de distance via l'effet de parallaxe auditive qu'ils sont ensuite capables de mettre correctement en relation avec l'espace réel.

Les erreurs relatives signées de localisation réalisées par les sujets lors de l'expérimentation ont aussi

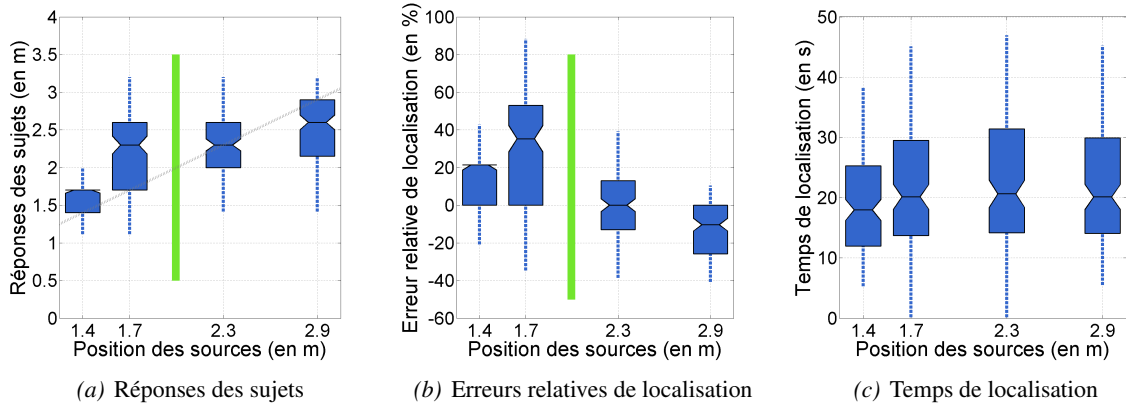


FIG. 7.19 – Résultats de l'expérience de restitution de l'effet de parallaxe. Sur la figure 7.19(a), la ligne pointillée grise indique la droite d'équation $y = x$. Sur les figures 7.19(a) et 7.19(b), le rectangle vert symbolise le SMART-I², qui est situé à 2 m de l'origine du repère.

été analysées à l'aide d'une ANOVA à mesures répétées prenant les positions des sources comme facteur. Les résultats obtenus indiquent un effet très significatif de la position de la source sur les réponses des sujets ($F(3) = 7.7$ et $p < 10^{-3}$). Un test post-hoc (t -test avec correction de Bonferroni) révèle à nouveau l'existence des 3 groupes de sources évoquées précédemment. Les erreurs réalisées par les sujets dépendent donc significativement de la distance à laquelle se situent les sources. En particulier, les distances des sources situées à 1.4 m et 1.7 m sont surestimées, celle de la source située à 2.3 m est correctement estimée, et celle de la source la plus lointaine (2.9 m) est sous-estimée (voir la figure 7.19(b)).

Les temps de localisation mis par les sujets pour estimer les distances des sources qui leur sont proposées ont enfin été analysées à l'aide d'une ANOVA à mesures répétées prenant les positions des sources comme facteur. Les résultats obtenus ne font ressortir aucun effet significatif de la position de la source sur les temps de localisation des sujets ($F(3) = 0.34$ et $p = 0.8$). Le temps que mettent les sujets à estimer la distance à laquelle se trouve les sources ne semble donc pas dépendre de la distance à laquelle elles se trouvent (voir la figure 7.19(c)).

Résultats de l'évaluation objective

Au cours de l'expérience, il a été demandé aux sujets de se déplacer dans la zone de rendu qui s'étend de l'origine du repère jusqu'à une position extrême située à $x = 1.5$ m et $y = 0$ m, comme le montre la figure 7.11. Au point de départ, les biais et variabilités correspondant aux ITD et ILD sont quasiment indépendants de la distance à laquelle se situe la source virtuelle et ne sont donc pas présentés ici. Les biais d'ITD sont de l'ordre de 0 ± 0.05 ms pour le système optimal de WFS et pour le SMART-I². Les biais d'ILD sont de 0 ± 0.5 dB pour le système optimal de WFS et de -1.5 ± 0.5 dB pour le SMART-I².

Les erreurs commises sur les indices de localisation à la position extrême ($x = 1.5$ m, $y = 0$ m sur la figure 7.11) sont représentés à la figure 7.20. Un biais positif ($B_{ITD} \simeq 0.3$ ms et $B_{ILD} \simeq 7$ dB) existe pour les sources sonores situées devant le LaMAP (1.4 m et 1.7 m) synthétisées à l'aide du SMART-I². Ce biais n'apparaît pas pour le système optimal de WFS. Ce biais, présent de façon cohérente à la fois pour

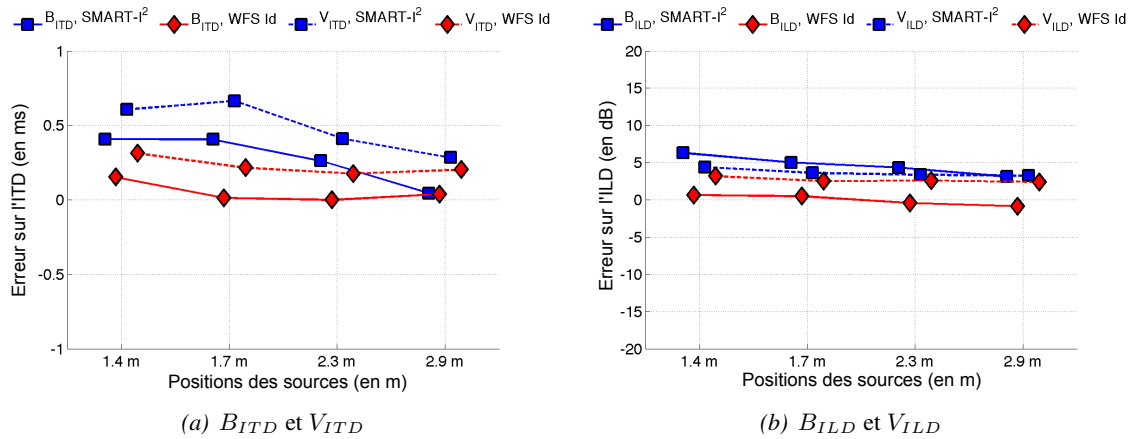


FIG. 7.20 – B_{ITD} et V_{ITD} (à gauche), B_{ILD} et V_{ILD} (à droite) déterminés à la position située à l’extrême droite de la zone de rendu pour le SMART-I² et le dispositif optimal pour la WFS. Les échelles verticales choisies sont représentatives des ITD et ILD attendus par le système auditif humain.

l’ITD et pour l’ILD, peut être responsable d’une partie de la sur-estimation des distances perçues observée dans l’expérience perceptive (voir figure 7.19(a)).

Discussion

Lors de l’expérience d’évaluation de la restitution du parallaxe auditive, les sujets ont globalement trouvé la tâche de localisation difficile à réaliser, ce qui n’était pas le cas pour l’expérience concernant l’azimut. Les sujets ont été assez perturbés par le fait que tous les stimuli avaient le même niveau sonore au point de départ et par l’absence d’effet de salle. Pour pallier ces problèmes, les sujets ont alors exploré la zone de rendu durant les premiers essais pour déterminer la position d’où il pouvait réaliser la meilleure estimation de distance, *i.e.* la position située à l’extrême droite de la zone de rendu. Après cette phase exploratoire, 80% des sujets se sont dirigés directement à cette position pour faire leur estimation. Il est donc pertinent de prendre en compte les résultats issus de l’analyse objective pour cette position particulière dans l’analyse des résultats obtenus ici.

De façon générale, les résultats de l’évaluation subjective font apparaître que les sujets arrivent à percevoir, dans une certaine mesure, les distances restituées par le SMART-I² grâce à l’effet de parallaxe auditive. Cela n’est pourtant pas évident, étant donné que la parallaxe auditive n’est pas aussi informative que les autres indices auditifs permettant d’estimer la distance [Speigle & Loomis 1993]. Mais ces résultats révèlent aussi une compression systématique des distances perçues. Ce phénomène de compression peut avoir différentes origines :

- Les résultats de l’analyse objective montrent que le rendu sonore spatialisé créé par le SMART-I² produit des erreurs systématiques des indices de localisation auditifs. Le système de rendu sonore spatialisé peut donc être à l’origine de la compression observée.
- La perception de la distance auditive basée sur des sources réelles fait apparaître un phénomène de compression similaire à celui observé ici (voir par exemple la figure 7.6). Dans une synthèse bibliographique sur le sujet, Zahorik et al. [2005] mettent en évidence le fait que les distances auditives

perçues sont naturellement sur-estimées pour des distances inférieures à $\simeq 2$ m et sous-estimées pour des distances supérieures, ce qui est bien le cas ici. Le phénomène de compression observé ici est donc peut être pour partie dû au système perceptif humain.

- Un phénomène de compression perceptive systématique des distances existe aussi dans les environnements visuels virtuels immersifs utilisant de grands écrans [Interrante et al. 2008, Grechkin et al. 2010]. Le rendu visuel 3D auquel est associé le rendu audio spatialisé est donc lui aussi susceptible d'induire une compression des distances.

Il est cependant délicat d'identifier la cause de la compression des distances observée ici, étant donné que le protocole expérimental utilisé n'est pas assez générique, et ne permet donc pas facilement de comparaisons avec d'autres études.

7.4.5 Effet de l'apprentissage

Le dernier point qu'il est possible d'évaluer est l'importance de l'ordre dans lequel ont été effectuées les deux expériences présentées précédemment. Pour analyser ce point, les erreurs issues des expériences de précision de localisation en azimuth et de restitution du parallaxe auditive ont été divisées en deux paquets : d'un côté le groupe AP qui a d'abord effectué la localisation en azimuth puis expérimenté l'effet de parallaxe, de l'autre le groupe PA qui a effectué ces tâches dans l'ordre inverse. Les résultats de ces deux groupes en termes d'erreurs de localisation en azimuth et en terme d'erreurs relatives d'estimation de la distance sont présentés sur la figure 7.21.

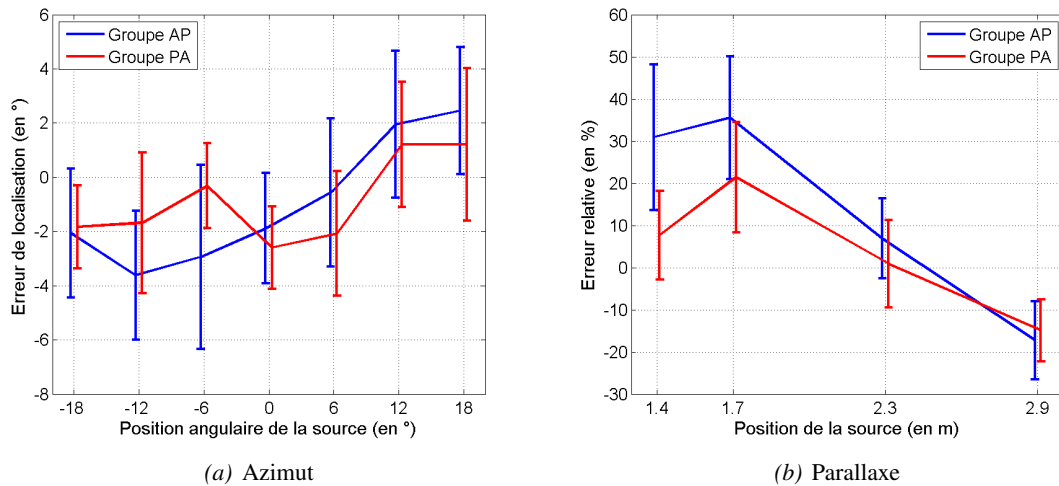


FIG. 7.21 – Influence de l'ordre dans lequel les expériences relatives à l'azimut et à la parallaxe ont été effectuée. Les lignes horizontales relient les valeurs moyennes et les traits verticaux indiquent les écart-types associés.

Les erreurs issues de l'expérience de localisation en azimuth ont été analysées à l'aide d'une ANOVA avec deux facteurs : le groupe d'appartenance du sujet et la position des sources en azimuth. Les résultats obtenus indiquent que le groupe n'est pas un facteur significatif ($F(1, 6) = 0.1$ et $p = 0.74$), que la position des sources en azimuth est un facteur très significatif ($F(6, 1) = 31$ et $p < 10^{-4}$), et qu'il existe une interaction

significative entre le groupe et la position angulaire des sources ($F(6, 6) = 6$ et $p < 10^{-4}$). Les sujets du groupe AP n'ont donc pas réalisé moins d'erreurs lors de la tâche de localisation en azimuth que ceux du groupe PA, mais, comme le montre la figure 7.21(a), ces erreurs sont distribuées différemment selon la position angulaire des sources (les sujets du groupe PA présentent moins de dilatation de l'azimut que ceux du groupe AP).

Les erreurs relatives issues de l'expérience liée à l'effet de parallaxe auditive ont aussi été analysées à l'aide d'une ANOVA avec deux facteurs : le groupe d'appartenance du sujet et la position des sources. Les résultats obtenus indiquent que le groupe est un facteur significatif ($F(1, 3) = 17$ et $p < 10^{-4}$), que la position des sources est un facteur très significatif ($F(3, 1) = 112$ et $p < 10^{-4}$) et qu'il existe une interaction significative entre le groupe et la position des sources ($F(3, 3) = 4.9$ et $p = 0.002$). Comme le montre la figure 7.21(b), les sujets du groupe PA réalisent donc moins d'erreurs lors de la tâche d'estimation de distance en utilisant l'effet de parallaxe que ceux du groupe AP, particulièrement en regard des deux sources les plus proches (1.4 m et 1.7 m).

En conclusion, il est possible de dire qu'un apprentissage des relations spatiales existant au sein du monde virtuel et du lien entre les mondes réel et virtuel a bien eu lieu au cours des deux expériences présentées ici. Ainsi, le groupe PA, peut être parce qu'il bénéficie d'abord d'une phase expérimentale plus libre (expérience sur la restitution de l'effet de parallaxe), semble mieux appréhender l'espace virtuel que le groupe AP.

7.5 Conclusion

L'objectif de ce chapitre est de déterminer dans quelle mesure le SMART-I² permet à un utilisateur, plongé dans l'environnement de réalité virtuelle qu'il restitue, d'utiliser ses capacités perceptives audio-visuelles et ses capacités motrices dans leur intégralité pour comprendre l'espace virtuel et se déplacer en son sein.

La façon dont l'espace est perçu au travers des modalités auditive et visuelle a d'abord été rappelée, puis les conditions perceptives requises pour que les flux spatio-temporels d'informations provenant des modalités auditive et visuelle fusionnent et donnent naissance à un unique flux audio-visuel ont été décrites. Deux expériences, ayant pour but d'estimer la précision de localisation en azimuth, d'étudier la capacité de restitution de l'effet de parallaxe auditive, ainsi que d'évaluer la cohérence perceptive du rendu audio-visuel proposé par le SMART-I² ont ensuite été présentées.

En ce qui concerne la précision de localisation en azimuth offerte par le système, il est possible de conclure que le SMART-I² génère un rendu audio spatialisé permettant une localisation en azimuth aussi précise, mais légèrement plus variable, que dans la réalité où qu'avec un système de WFS conventionnel. D'autre part, l'analyse objective suggère que la qualité de localisation observée ici est identique dans toute la zone de restitution, démontrant ainsi que la WFS ne souffre pas de problèmes de *sweet spot*.

À propos de la capacité du SMART-I² à restituer l'effet de parallaxe auditive, les résultats montrent qu'effectivement le système est en mesure de procurer à ses utilisateurs une sensation de distance réaliste par ce biais. Cependant les résultats obtenus révèlent aussi un phénomène de compression perceptive de la distance ayant potentiellement plusieurs origines : le dispositif de rendu sonore spatialisé, le système auditif humain, ou alors le rendu visuel 3D.

Enfin, la dernière conclusion que l'on peut tirer de cette étude est que les sujets qui ont participé à cette étude ont été capable de percevoir les sources sonores dans l'espace virtuel et ont été en mesure de transporter avec succès cette information auditive dans la modalité visuelle lors du report. Cela suggère donc que l'environnement virtuel proposé par le SMART-I² présente suffisamment de cohérence audio-visuelle pour que les sujets soient à même d'intégrer les flux auditif et visuel qui leur sont présentés et d'agir dans cet espace. Les effets d'apprentissage observés semblent confirmer ce point. Il semble donc que le SMART-I² réunisse les conditions requises (voir section 7.3) pour permettre une bonne intégration perceptive des flux auditifs et visuels.

Une perspective intéressante ouverte par ce chapitre concerne la simulation de l'élévation par le système de rendu sonore spatialisé. En effet, la "*Wave Field Synthesis*" étant basée sur un banc de haut-parleurs répartis sur une ligne horizontale et disposés à hauteur de tête (voir section 2.5.3), l'élévation n'est pas restituée par le dispositif dans sa version actuelle. Avant de se lancer dans des considérations d'ordre technologique, la première question à laquelle il faut répondre est : "*Est-il, dans des applications audio-visuelles de réalité virtuelle, nécessaire de restituer l'élévation des objets sonores virtuels ?*". Les analyses bibliographiques des sections 7.3.1 et 2.3 suggèrent que c'est possiblement le cas, mais que cela mérite d'être étudié de façon rigoureuse. Si le rendu de l'élévation des objets sonores s'avère nécessaire, plusieurs solutions technologique peuvent alors être envisagées. Une première solution est de coller au dos des La-MAPs plusieurs lignes d'excitateurs électrodynamiques (voir section 2.5.4) et de sélectionner la (ou les) lignes à piloter en fonction de l'élévation de la source. Une seconde solution consiste à simuler l'élévation des objets sonores virtuels par un filtrage fréquentiel similaire à celui qui serait effectué par le pavillon des oreilles à partir de l'analyse d'une base de donnée d'HRTFs ⁷, comme le font par exemple Lopez et al. [2010].

7. On se référera à la section 2.4.2 pour quelques mots concernant les HRTFs ou bien à [Nicol 2010] pour de plus amples détails.

Chapitre 8

Transparence de l'interface de restitution et perception de l'espace virtuel

Sommaire

8.1	Introduction	205
8.2	Distance égocentrique audio, visuelle et audio-visuelle perçue en environnement virtuel par des sujets mobiles	206
8.2.1	Introduction	206
8.2.2	Méthode	209
8.2.3	Analyse des résultats	213
8.2.4	Discussion	219
8.2.5	Conclusion	225
8.3	Discussion	226
8.3.1	Retour sur l'intégration audio-visuelle en distance	226
8.3.2	Limites de la procédure de triangulation (étude préliminaire)	228
8.3.3	Comparaison des deux expériences d'estimation de la distance audio	230
8.4	Conclusion	232

8.1 Introduction

Dans le chapitre précédent, la façon dont un utilisateur, plongé dans un environnement de réalité virtuelle synthétisé par le SMART-I², perçoit l'espace virtuel et est capable de se déplacer en son sein a été étudiée. Ni ses attentes ni sa connaissance a priori du dispositif et du monde virtuel n'ont été prises en compte dans cette étude. Il a cependant été montré dans la section 2.2 que dans les environnements de réalité virtuelle *perception* et *cognition* ne sont pas indépendantes l'une de l'autre. La *perception* désigne ici la façon dont l'environnement virtuel est compris en fonction de ce que l'on perçoit de lui. La *cognition* désigne la façon dont l'environnement virtuel est compris en fonction de ce que l'on sait de lui. Par exemple, une augmentation de l'*immersion* ne garantit pas forcément une augmentation de la sensation de *présence* perçue par les utilisateurs parce que l'on *sait* potentiellement que l'environnement qui nous est présenté est

virtuel. Le fait que la fusion audio-visuelle soit influencée par l'éventuel lien sémantique que l'on *sait* (ou non) exister entre les stimuli auditifs et visuels, ou encore le fait que la distance auditive ou visuelle perçue résulte en partie de la *connaissance* que l'on a de l'objet étudié, illustrent aussi ce lien.

Ce chapitre s'attache donc à comprendre comment la *perception* de l'environnement audio-visuel synthétisé par le SMART-I² est liée à la *connaissance* que l'on a du dispositif physique. Plus précisément, une expérience centrée sur l'influence de la *transparence* limitée du SMART-I² sur la perception de la distance et sur la sensation de *présence* est présentée à la section 8.2. La section 8.3 discute ensuite les résultats de cette expérience du point de vue de l'*intégration audio-visuelle* et les compare à ceux obtenus à la section 7.4.4.

8.2 Distance égocentrique audio, visuelle et audio-visuelle perçue en environnement virtuel par des sujets mobiles

Marc Rébillat, Xavier Boutillon, Étienne Corteel, Brian F.G. Katz

Cette section reproduit l'article intitulé "*Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments*" soumis au journal *ACM Transactions on Applied Perception* (ACM-TAP) le 20 Août 2011 [Rébillat et al. 2011a]. Une partie de cet article a fait l'objet d'une présentation lors du *Forum Acusticum 2011* ayant eu lieu à Aalborg [Rébillat et al. 2011b].

Abstract

A study on *audio*, *visual*, and *audio-visual* egocentric distance perception by moving subjects in *virtual* environments is presented. Audio-visual rendering is provided using tracked passive visual stereoscopy and acoustic wave field synthesis (WFS). Distances are estimated using indirect blind-walking (triangulation) under each rendering condition. Experimental results suggest that: i) distances perceived in the virtual environment are systematically overestimated for rendered distances closer than 2.3 m and underestimated for distances farther than 2.3 m, whatever the rendering condition is, ii) the distance underestimation observed in visual virtual environments is linked to the *presence* feeling for more than 75% of the subjects: the more present the subject feels, the less underestimated the distance is, iii) there is no difference in the perception of auditory distance in the action space (1.5 to 5 m) between real environments and virtual environments rendered by WFS, iv) the audio-visual virtual environment is found to be scaled according to the physical position of the audio-visual virtual reality setup in the real environment, which is thus proposed to act as an anchor between both worlds.

8.2.1 Introduction

Virtual reality (VR) systems aim at providing subjects with a virtual world where they would behave and learn as if they were in the real world [Brooks 1999]. Audio-visual (AV) VR-systems that combine large immersive screens and many loudspeakers have been developed to provide subjects with a virtual space coherently merging holophonic spatial audio and 3D visual renderings [Faria et al. 2005b, Springer et al. 2006, Rébillat et al. 2008]. The term "*holophonic spatial audio*" stands here for technologies such

as wave-field synthesis (WFS) [Berkhout et al. 1993], Ambisonics [Gerzon 1985], or others [Komiya et al. 1991], that recreate physically the same sound field that a real sound source would have radiated and thus provide subjects with a natural spatialized sound rendering. Such AV VR-systems are very appealing because they are minimally intrusive (no headphones needed, only lightweight glasses) and allow subjects to move freely in the rendering area while always having a correct AV perspective. With the emergence of these multimodal systems arises the question of the correct perception of the virtual space by moving subjects, and more specifically of rendered distances within it [Loomis & Knapp 2003, Interrante et al. 2008].

Measurement protocols for the estimation of perceived egocentric distance

Because distance perception is a cognitive task, measurement protocols are needed to estimate perceived absolute egocentric distances. Existing measurement protocols can be divided into three main classes [Klein et al. 2009, Grechkin et al. 2010]: verbal estimations, perceptually directed actions, and imagined actions. In *verbal estimation* protocols, subjects assess the perceived distance in terms of familiar units, such as meters. In *perceptually directed action* protocols, an object is presented to the subject who then has to perform an action, such as blind-walking, without perceiving the object any more. In *imagined action* protocols, the action is imagined instead of being performed and response times are used to infer the results of the action. The advantage of *perceptually directed actions* is that they lead to distance estimations that are more accurate and less variable than distance estimations provided by verbal reports [Fukushima et al. 1997, Loomis et al. 1998, Russell & Schneider 2006, Andre & Rogers 2006]. Moreover, using *perceptually directed actions*, estimated distances can be directly inferred from actions whereas a potential systematic bias exists in distances estimated using *imagined action* protocols due to the conversion of a directly measured value of time into an indirect measure of estimated distance [Grechkin et al. 2010]. *Perceptually directed actions* have thus been preferred in the present study.

Among perceptually directed actions, *direct* blind-walking and *indirect* blind-walking (triangulation) are two possible alternatives which both lead to accurate distance estimations [Fukushima et al. 1997, Loomis et al. 1998]. Due to physical spatial constraints imposed by the presence of large screens and many loud-speakers, only indirect blind walking (triangulation) is possible in the kind of AV VR-systems under study here [Klein et al. 2009]. An advantage of the triangulation measurement protocol is that it is applicable to the measurement of audio, visual, and audio-visual perceived absolute egocentric distances without any need to adapt the procedure to each different modality.

Perceived distance in visual VR-systems based on large immersive screens

In classical *visual* VR-systems, such as head-mounted displays (HMD), perceived visual distances have been observed to be systematically underestimated [Loomis & Knapp 2003, Interrante et al. 2008]. This is not the case in the real world [Wiest & Bell 1985]. VR-systems based on large immersive screens were thought to offer a better distance perception [Plumert et al. 2005]. Studies focusing on visual distance perception in virtual environments rendered by large immersive screens has found that visual distances were underestimated using these systems, exactly as in HMD systems [Armbruster et al. 2008, Naceri et al. 2009, Klein et al. 2009, Grechkin et al. 2010, Alexandrova et al. 2010]. The reasons for this underestimation

are still not well understood. Among potential factors, it seems that a poor feeling of *presence* (sensation of being “*in the scene*”, see [Schuemie et al. 2001]) experienced by subjects immersed in the virtual environments may contribute to such underestimations [Interrante et al. 2008].

It may thus be expected that because they potentially provide a higher degree of *presence* [Bormann 2005] than visual-only VR-systems based on large immersive screens, AV VR-systems combining holophonic spatial audio and large immersive screens will lead to a lesser degree of distance underestimation.

Perceived distance in *audio* VR-systems based on holophonic sound rendering

In the *real* world, it is well established that near-auditory distances (< 2 m) are overestimated whereas far-auditory distances (> 2 m) are underestimated (see [Zahorik et al. 2005] for a review). Much less is known regarding *auditory* distance perception in virtual auditory systems based on holophonic spatial audio. In [Noguès et al. 2003, Corteel 2004, Rébillat et al. 2008], it was experimentally shown that holophonic spatial sound renderings can effectively be used to render distances between static sources and moving subjects and that perceived distances are compressed with respect to rendered distances. When subjects are static, studies [Komiya et al. 1991, Kearney et al. 2010] show that performances in an holophonic audio virtual environment match well with real world performances in terms of distance perception. In another study by [Wittek et al. 2004], it is contradictorily underlined that subjects have to move in order to receive the full benefits of holophony for distances closer than 2 m.

At this point, it is not well established to what extent distances simulated by holophonic spatial sound renderings are perceived by moving subjects as they would be perceived in the real world.

Perceived distance in *audio-visual* VR-systems

As stated in Sec. 8.2.1 regarding *virtual* environments, all perceived visual distances appear to be underestimated, near-auditory distances (< 2 m) to be overestimated, and far-auditory distances (> 2 m) to be underestimated. In *virtual* environments audio and visual perceived distances are thus *a priori* inconsistent for a given rendered distance. Some efforts have thus been done to study how audio and visual distance cues are merged together in virtual environments [Côté et al. 2011]. Obtained results suggest that static subjects perceived AV distances similarly to visual distances.

Since subjects are rarely static when performing tasks in VR environments, it is important to study whether perceived AV distances are also similar to visual distances for moving subjects taking benefit of static and dynamic AV distances cues.

Impact of the audio-visual rendering setup on distance perception

To provide subjects with a virtual world where they would behave as if they were in the real world, VR-systems should be fully *transparent* to subjects. Transparency is closely linked to the notion of *immersion*, understood here as “*the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of a human participant*” [Slater & Wilbur 1997]. However, AV VR-systems are not perfect and suffer from some drawbacks that limit the degree of *immersion*. It is thus important to know if this limitation has an influence on the AV virtual world perceived

by subjects, and what kind of spatial link do AV VR-systems create between the real world and the virtual world.

Objectives

In this paper a study of *audio* (A), *visual* (V), and *audio-visual* (AV) egocentric distance perception in the action space (1.5 m to 5 m) by moving subjects in *virtual* environments is presented. AV rendering is provided via the SMART-I² platform (Spatial Multi-user Audio-visual Real-Time Interactive Interface) [Rébillat et al. 2008, 2009] using tracked passive visual stereoscopy and acoustic wave field synthesis (WFS). This AV VR-system allows subjects to move freely in the rendering area while always having a correct AV perspective. Distances are estimated by means of perceptually directed action -indirect blind walking: triangulation- under A, V, and AV conditions. This experiment aims at investigating the following points:

1. Testing the hypothesis that because they potentially provide a high degree of *presence* AV VR-systems based on holophonic spatial audio lead to less visual distance underestimation than visual-only VR-systems.
2. Evaluating to what extent distances simulated by holophonic spatial sound renderings are perceived by moving subjects as they would be perceived in the real world.
3. Studying if distances perceived by moving subjects in an audio-visual virtual world are similar to distances perceived by moving subjects in a visual-only or in an audio-only virtual world.
4. Assessing whether the fact that the AV rendering system is not fully transparent induces a spatial link between the real and virtual worlds.

8.2.2 Method

Experimental design

In the present experiment, five distances in the subject's *action space*, *i.e.* the space where one “*moves quickly, talks, and if needed can throw something to a compatriot or at an animal*” [Cutting 1997], were tested: 1.5 m, 2 m, 2.7 m, 3.5 m, 5 m.

A total of 20 volunteers between 24 and 45 years old participated in the experiment (17 men). All subjects had self-reported normal vision (possibly corrected) and normal hearing. Each subject had to estimate the five distances four times under each rendering condition. They performed three sessions of 20 iterations each after a training phase of two iterations under each rendering condition. In the training phase rendered distances were 3 m and 7 m. Subjects took pauses between sessions and the entire experiment lasted approximately one hour. The session order was balanced between the six possible permutations of the three conditions.

The chosen experimental design was therefore a repeated-measures design with two factors: the rendered distance d_r (five levels, fixed factor) and the rendering condition (three levels, fixed factor). The dependant variables are the perceived distance d_p , the relative error between rendered and perceived distances ϵ_p , the time t_{xp} spent in the exploration phase (see Sec. 8.2.2), and the exploration path length l_{xp} . Tab. 8.1 summarises the chosen experimental design.

<i>Independent variables</i>			<i>Dependent variables</i>	
Subject	20	Random variable	Perceived distance	d_p (m)
Rendering condition	3	A, V, AV	Relative error	ϵ_p (%), see Eq. (8.2)
Rendered distance d_r (m)	5	1.5, 2, 2.7, 3.5, 5	Exploration time	t_{XP} (s)
Repetition	4	1, 2, 3, 4	Exploration path	l_{XP} (m)

Table 8.1: Independent and dependent variables used in the experimental protocol

Experimental setup

Experiments were conducted in the AV virtual environment produced by the SMART-I² platform [Rébillat et al. 2008, 2009]. In this system, front-projection screens and loudspeakers are integrated together to form large flat multi-channel loudspeakers also called *Large Multi-Actuator Panels* (LaMAPs). The rendering screens consist of two LaMAPs (2 m × 2.6 m, each supporting 12 loudspeakers) forming a corner (see Fig. 8.1). The report interface used in the present experiment was a wiimote.

Visual rendering was produced using tracked passive stereoscopy rendered at 80 frames per second with a resolution of 1280 × 960 pixels on each screen. At the starting position (the black «×» in Figs. 8.2(a) and 8.2(b)), the horizontal field of view was approximately 150° and the vertical field of view approximately 70°. Since graphical resolution [Thompson et al. 2004, Kunz et al. 2009] and field of view [Creem-Regehr et al. 2005] have been shown to have no influence on *visual* distance perception, these experimental parameters should not influence the obtained experimental results.

Spatial audio rendering was realized by acoustic wave field synthesis [Berkhout et al. 1993]. This technology physically recreates the acoustic sound field corresponding to a virtual source at any given position in the horizontal plane, without the need for tracking. The inter-loudspeaker distance of 21 cm corresponds to an aliasing frequency $f_{al} = 1070 \pm 70$ Hz, up to which the sound field is spatially exactly reconstructed [Cortel 2004]. The perceptual effects of the aliasing frequency f_{al} are discussed in details in Sec. 8.2.4.

Furthermore, fine temporal and spatial calibration has been done to ensure that the audio and visual renderings are fully coherent.

Audio, visual, and audio-visual stimuli

The *visual environment* consisted of an open, grassy field with a forest at 50 m (Fig. 8.1(a), trees are $\simeq 7$ m tall). The associated *audio environment* consists of the sound of wind in the trees accompanied by some distant bird songs (overall background level of 36 dBA). The audio environment was spatialized using 12 plane waves equally distributed in the horizontal frontal field of rendering (*i.e.* between -70° and 70° , see Fig. 8.2(a)). Environmental sound levels were adjusted to be slightly above the background noise produced by the video-projectors (background noise level of 34 dBA).

The chosen *visual target* object was a footless 3D loudspeaker, approximately spherical, with a diameter of $\simeq 30$ cm (Fig. 8.1(a)). The stand was removed to avoid window violations when the object was displayed in front of the screen. The floating loudspeaker was positioned at a height of 1.6 m and shadows were displayed.

The associated *audio target* object was a 4 kHz low-pass filtered white noise with a 15 Hz amplitude modulation. White noise has been chosen in order to have a wide spectral content and to allow subjects

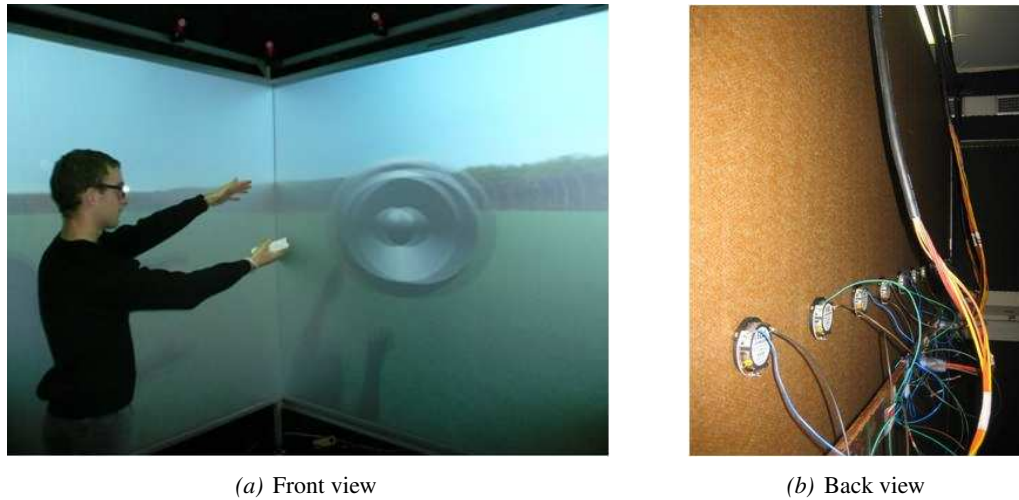


Figure 8.1: *Left*: a subject and an audio-visual object in the virtual world provided by the SMART-I². Visual rendering is projected on the front faces of the two LaMAPs which form a corner. *Right*: electro-dynamical exciters are glued on the back of each LaMAP.

to rely on numerous audio localisation cues. The white noise was modulated in amplitude by a sine wave to produce attack transients that are also useful in sound localization [Blauert 1999]. The choice of the 4 kHz cut-off frequency is discussed in Sec. 8.2.4 regarding the aliasing frequency f_{al} . No room-effect (*i.e.* ground reflections) was included. The sound level of the audio object corresponds to a monopole emitting 78 dB(SPL) at 1 m, well above the environmental sound level at each of the tested distances.

Audio and *visual* objects were always displayed coherently, *i.e.* at the same spatial position. In addition, their *visual* size and *audio* level decreased naturally with distance. As the experimental design allowed subjects to move within the rendering area (see Sec. 8.2.2), they could rely on a large number of cues naturally available in the corresponding real environment for the estimation of distances, including dynamic cues. In particular, *motion parallax*, which denotes changes in the angular direction of a point source occasioned by the subject's translation is available. This cue has been shown to be useful for distance estimation using the visual [Beall et al. 1994, Nawrot & Stroyan 2009] or the auditory modality [Speigle & Loomis 1993, Porschmann & Storig 2009]. Another dynamic cue, the *acoustic* τ , which corresponds to the estimated time-to-collision for a constant velocity between the moving subject and the static source, can also be used [Ashmead et al. 1995, Porschmann & Storig 2009]. Available AV distance cues are summarized in Tab. 8.2.

Available cue	Modality	Class
Object size/level	A, V	Relative
Motion parallax	A, V	Absolute
Acoustic τ	A, V	Absolute
Binocular/binaural cues	A, V	Absolute
Height in the visual field	V	Absolute

Table 8.2: Available audio-visual cues.

The AV background environment was kept active in all the rendering conditions. In the *audio* condition, the spatialized sound corresponding to the virtual object was played while no image of the virtual object was shown. The only visual image consisted of the open, grassy field with a forest in the background. In the *visual* condition, the 3D image of the virtual object was displayed with no corresponding sound. The only audio signal consisted of the sound of wind in the trees accompanied by some bird songs. In the *audio-visual* condition, the spatialized sound of the virtual object was rendered with its corresponding 3D image and the AV environment.

Experimental task

Distance estimation was performed here in two phases: an *exploration* phase, see Fig. 8.2(a), and a *triangulation* phase, see Fig. 8.2(b). Subjects began each iteration at the *start position*, indicated by a black «X» in Figs. 8.2(a) and 8.2(b). Virtual objects were displayed on the right screen of the SMART-I² platform, see Fig. 8.1(a), and their rendered distance d_r was relative to the start position, see Fig. 8.2(a).

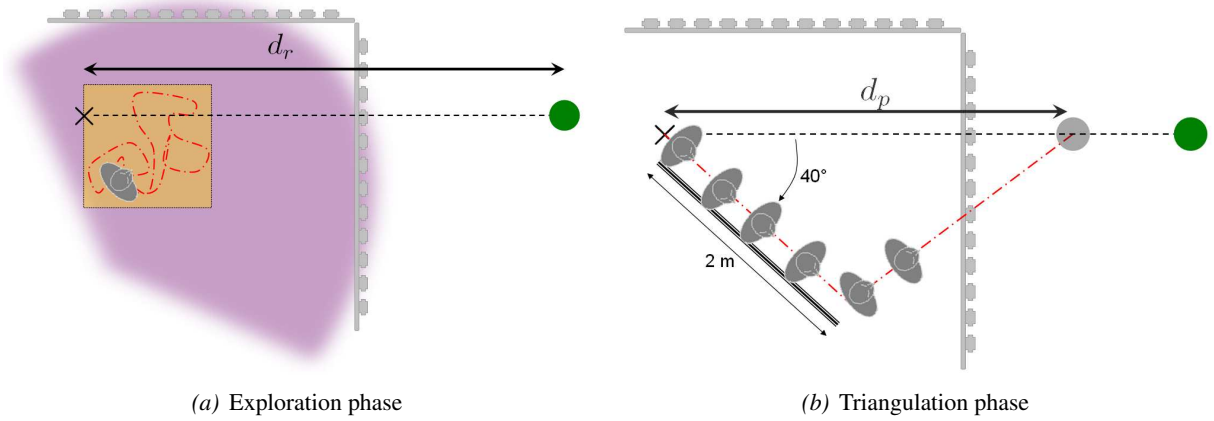


Figure 8.2: Exploration and triangulation phases. *Start position*: black «X». *Virtual object*: green disk placed at a distance d_r from the *start position*. In Fig. 8.2(a), the *exploration area* is represented by the yellow rectangle and the dotted red line (— · —) indicates a hypothetical exploration trajectory performed by the subject. The purple cloud schematically represents the audio surrounding environment. In Fig. 8.2(b), the *guide* is shown as a thick plain black line. The dotted red line (— · —) indicates a hypothetical trajectory performed by the subject. *Perceived object*: grey disk placed at the estimated perceived distance d_p from the *start position*.

In the *exploration* phase, subjects move around in the *exploration area* which was a rectangle of $1 \times 0.8 \text{ m}^2$. During this phase, subjects were instructed to move in the *exploration area* in order to acquire “a good mental representation of the virtual object and of its environment”. Before starting this phase, subjects had to indicate that they were ready to perform the exploration phase by pressing a wiimote button. A typical *exploration* phase is depicted in Fig. 8.2(a).

Once a “a good mental representation” has been acquired, subjects were asked to press a wiimote button to indicate that they were ready for the *triangulation phase*. At this point, the perceived distance d_p was estimated by means of triangulated blind walking, as depicted in Fig. 8.2(b). Subjects closed their eyes, made a 40° right-turn and walked blindly for an imposed distance of $\simeq 2 \text{ m}$, following a handrail guide which was included to help during blind-walking. Subjects stopped at the end of the guide, turned in the

direction where they thought the object was, and took a step forward in the direction of the source position. Subjects had been told that the perceived distance was to be calculated according to this step. They then indicated that they had completed the triangulation by pressing a wiimote button. Afterwards, they could open their eyes and go back to the *start position* for the next iteration. The experimental protocol was fully automated, with the subjects being observed remotely so as not to disturb the sense of presence.

Post-session questionnaire

Post-session questionnaires are the most common tools to obtain the subjects feedback about their experience in the virtual environment [Insko 2003]. In the present experiment, subjects were asked to complete a 15-item questionnaire at the end of each of the three sessions of the whole experiment (A, V, AV). The goal of this questionnaire was to evaluate the feeling of *presence*, the perceived *spatial quality*, and the *task difficulty* that subjects experienced during each session. This questionnaire was built mainly by adapting statements taken from [Bormann 2005] and [Armbruster et al. 2008] that were translated into French. Statements were rated on a 7-point Likert scale ranging from -3 to 3 with two anchors. The statements are reported in Tab. 8.3.

Q1 ^{a†} :	I had the feeling of locating a real object.
Q2 ^{a†} :	I had the feeling of looking at TV instead of really being in an outdoor environment.
Q3 ^{a†} :	The virtual environment became real for me and I forgot the real environment.
Q4 ^{a†} :	I remember the virtual environment more as a place where I have been than as a computer generated image I have seen.
Q5 ^b :	My experience in the virtual environment was coherent.
Q6 ^{a*} :	I had the impression that I could touch the virtual objects.
Q7 ^{b*} :	I clearly perceived the virtual world and the virtual objects.
Q8 ^{b*} :	The audio-visual rendering quality was satisfying.
Q9 ^a :	I felt present in the virtual world.
Q10 ^{a†} :	I felt surrounded by the virtual world.
Q11 ^{b*} :	I found the depth sensation to be realistic.
Q12 ^b :	I found the distances and sizes or intensities of the virtual objects coherent.
Q13 ^b :	I had a good 3D vision of the virtual world and of the virtual objects.
Q14 ^c :	I could easily estimate the distances of the virtual objects.
Q15 ^c :	I trust in my estimations of the distances of the virtual objects.

Table 8.3: Post-session questionnaire.

^a:Statements used to compute the *presence* score: P-score (see Sec. 8.2.3). ^b:Statements used to compute the *spatial quality* score: SQ-score (see Sec. 8.2.3). ^c:Statements used to compute the *task difficulty* score: TD-score (see Sec. 8.2.3). [†]:Statements adapted from [Bormann 2005]. ^{*}:Statements adapted from [Armbruster et al. 2008].

8.2.3 Analysis of results

Because differences larger than twice the standard deviation were observed on the mean estimated distances for 3 subjects relative to the mean estimated distances of all subjects, data from these 3 subjects were removed from further analysis. Only results from the 17 remaining subjects (14 men) are shown in this section.

Extraction of the dependant variables

This section explains how the different dependent variables (d_r , ϵ_p , t_{XP} , l_{XP} , defined in Tab. 8.1) were derived from the experimental data. The position of the head of the subject (central point between the eyes) and corresponding times are recorded for each iteration during both the *exploration* and *triangulation* phases at 100 Hz. An illustrative example of real *exploration* and *triangulation* trajectories is shown in Fig. 8.3.

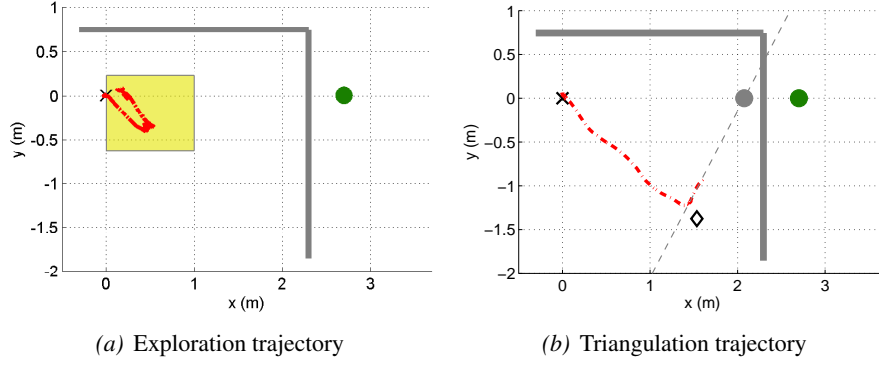


Figure 8.3: Example of *exploration* and *triangulation* trajectories (subject 18, iteration 3, *audio* condition, $d_r = 2.7$ m and $d_p = 2.07$ m with a 95% confidence interval of [1.99 m, 2.16 m]). *Exploration area*: yellow rectangle. *Beginning position and origin of the (x,y) coordinates*: black «x». *End of the guide*: black diamond. *Virtual object*: green disk. *Perceived object*: grey disk. *Subject trajectories*: dotted red line (— · —).

The duration of the *exploration* phase t_{XP} was obtained by measuring the time between the subject's action on the wiimote button for *ready for a new trial* and for *ready for the triangulation phase*, as explained in Sec. 8.2.2. The exploration path length l_{XP} was calculated by using the head positions recorded during t_{XP} .

The exploration path length walked during the exploration phase l_{XP} can be divided into portion walked in the direction of the source l_{XP}^P and the portion walked in the orthogonal direction l_{XP}^O . To be comparable, these two paths were normalized by the maximum physical path lengths in each direction, which are here the sides of the exploration area (*i.e.* $s_P = 1$ m and $s_O = 0.8$ m). The dependent variable $P = l_{XP}^P / s_P$ (respectively $O = l_{XP}^O / s_O$) denotes the number of times has walked by the subject the length of the exploration area parallel (respectively orthogonal) to the source direction.

Perceived distances d_p were estimated from the *triangulation* trajectory as follows. A line ($y = ax + b$) is fitted to the trajectory points during the forward step (125 ± 60 points have been used for the fit). The estimated perceived distance is given by Eq. (8.1), see Fig. 8.3(b):

$$d_p = -\frac{b}{a} \quad (8.1)$$

The relative 95%-confidence intervals on the estimated distance are deduced from the 95%-confidence intervals of the linear fit regression coefficients a and b for each iteration (*regress* function in Matlab). Accross all iterations, subjects, and distances, the median of the 95%-confidence intervals for the relative

distances, *i.e.* for d_p/d_r , estimated using the triangulation trajectory is $\pm 9.9\%$. The triangulation procedure and the associated data treatment thus provide a reliable estimation of the perceived distances.

The relative error ϵ_p made by subjects when estimating distances using the triangulation procedure is:

$$\epsilon_p = 100 \times \frac{d_p - d_r}{d_r} \quad (8.2)$$

Exploration phase

In this section, the influence of the rendered distance d_r and of the condition (A,V,AV) on the time t_{XP} and on the path length l_{XP} respectively spent and walked during the *exploration phase* are analyzed. Results of the analysis are shown in Fig. 8.4.

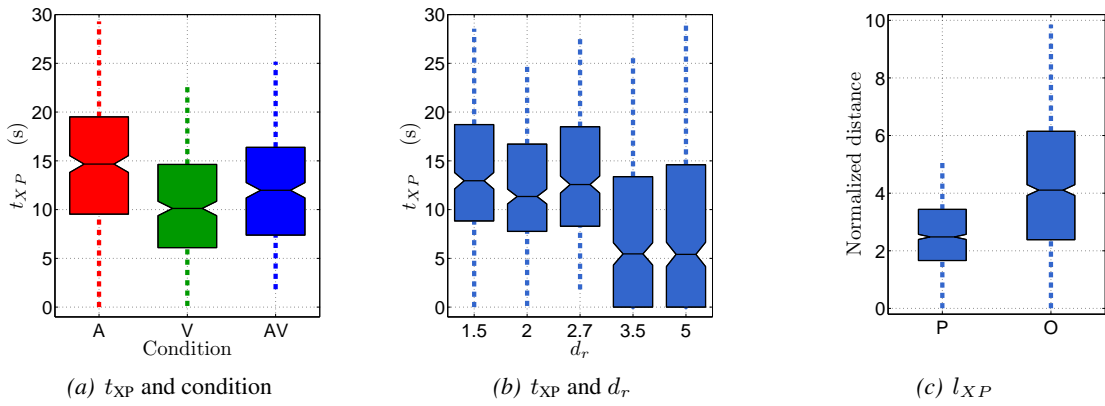


Figure 8.4: Exploration time (t_{XP}) as a function of the condition and the rendered distance (d_r), and comparison of the normalized exploration path lengths in the direction of the virtual object (P) and in the perpendicular direction (O). On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points (outliers not considered). Points are drawn as outliers if they are larger than $q_3 + 1.5(q_3 - q_1)$ or smaller than $q_1 - 1.5(q_3 - q_1)$, where q_1 and q_3 are the 25th and 75th percentiles, respectively. Notches denote comparison intervals. Two medians are significantly different at the 5% significance level if their intervals do not overlap. Interval endpoints are the extremes of the notches or the centers of the triangular markers.

A two-way repeated-measures analysis of variance (ANOVA) performed on the *exploration time* t_{XP} with *condition* (A,V,AV) and *rendered distance* d_r as within-subject factors shows that *condition* is highly significant ($F(2, 64) = 7.7$ and $p = 0.002$), that *rendered distance* almost reaches significance ($F(4, 64) = 2.4$ and $p = 0.056$) and that there is no interaction effect between *condition* and *rendered distance* d_r ($F(8, 64) = 1.2$ and $p = 0.31$). Post-hoc tests (*t-test* with Bonferroni correction) reveal that *exploration times* for each *condition* are significantly different. Post-hoc tests performed on *rendered distances* (*t-test* with Bonferroni correction) reveal that *exploration times* for *rendered distances* $d_r = 1.5, 2$ or 2.7 m are significantly different from those obtained for *rendered distances* $d_r = 3.5$ and 5 m. Subjects spent more time in the *exploration phase* when estimating distances using the audio modality than when using the audio-visual modality. Furthermore subjects spent more time in the *exploration phase* when estimating distances using the audio-visual modality than when using only the visual modality. Moreover, significantly

less time was spent exploring the virtual environment for far rendered distances (3.5 m and 5 m) than for close distances (1.5 m, 2 m and 2.7 m).

A two-way repeated-measures ANOVA performed on the exploration path length l_{XP} with *condition* (A,V,AV) and *rendered distance* d_r as within-subject factors shows that *condition* is highly significant ($F(2, 64) = 5.5$ and $p = 0.009$), that *rendered distance* is not significant ($F(4, 64) = 0.51$ and $p > 0.7$) and that there is no interaction effect between *condition* and *rendered distance* d_r ($F(8, 64) = 0.38$ and $p > 0.9$). Post-hoc tests on *condition* (*t-test* with Bonferroni correction) show that *exploration path lengths* for the audio condition are significantly higher from those for the visual and the audio-visual conditions. The same holds for the normalized exploration path lengths in the direction to the virtual object (P) and in the orthogonal direction (O). A shorter distance was thus walked by subjects during the exploration phase when the visual modality was involved than when it was not the case.

A *t-test* has also been performed to compare the normalized exploration path lengths in the direction of the virtual object (P) and in the orthogonal direction (O). This analysis reveals that O is significantly greater than P ($F(988) = 29.4$ and $p < 10^{-5}$). Given a certain exploration area, subjects walked 1.6 times more in the direction perpendicular to the virtual object than in the direction parallel to the virtual object during the exploration phase.

Triangulation phase

In this section, the influence of the rendered distance d_r and of the rendering condition (A,V,AV) on the dependant variables, *i.e.* the perceived distances d_p and the relative errors ϵ_p , are analysed. Results are shown in Fig. 8.5.

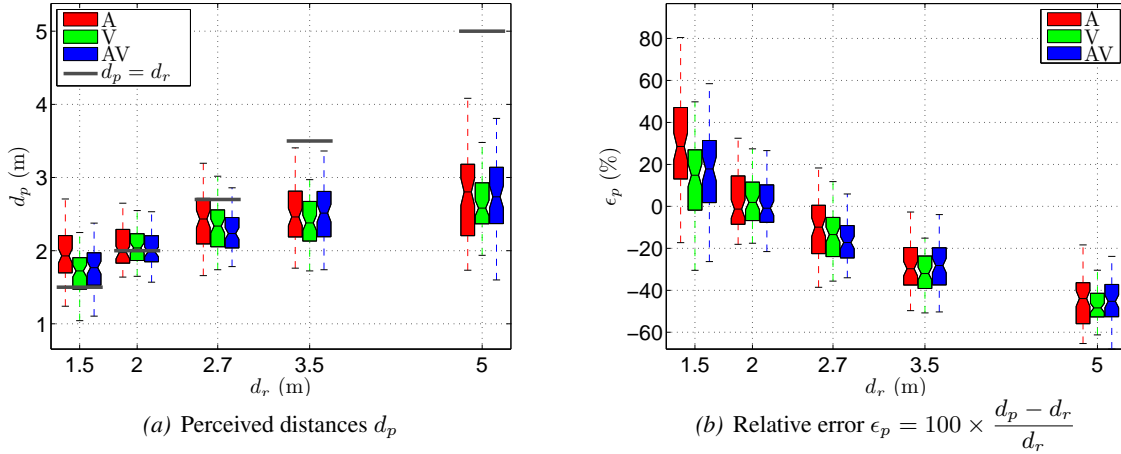


Figure 8.5: Perceived distances d_p and relative error ϵ_p as functions of the rendered distance d_r for each rendering condition. See the caption of Fig. 8.4 for explanations regarding plot. In Fig. 8.5(a), the grey horizontal lines represent a "perfect" subject with $d_p = d_r$.

The perceived distances d_p were analyzed using a repeated-measures two-way ANOVA with *rendered distance* d_r and *rendering condition* (A,V,AV) as factors. *Rendered distance* d_r is significant at the 5% level with $F(4, 64) = 93$ and $p < 10^{-6}$. *Rendering condition* is not significant at the 5% level as $F(2, 64) =$

1.96 and $p = 0.09$. No significant effect could be seen at this level between *rendered distance* d_r and *condition* since $F(8, 64) = 2$ and $p = 0.057$. As post-hoc tests, a series of Bonferroni corrected *t*-tests have been performed and all the *rendered distance* pairs have been found to be significantly different. The different distances d_r are thus correctly ordered and well recognized by the subjects, independently of the rendering condition. Interestingly, all the rendered distances are perceived consistently across the different available conditions.

The relative errors ϵ_p were analysed using a repeated-measures two-way ANOVA with the following within-subjects factors: *rendered distance* d_r and *rendering condition* (A,V,AV). *Rendered distance* d_r is significant at the 5% level with $F(4, 64) = 240$ and $p < 10^{-6}$. *Rendering condition* is only slightly significant at the 5% level as $F(2, 64) = 3.7$ and $p = 0.037$. The interaction effect between *rendered distance* and *condition* is also found to be significant with $F(8, 64) = 3.2$ and $p = 0.002$. Bonferroni corrected *t*-tests have been performed on d_r and all the *rendered distance* pairs have been found to be significantly different. Bonferroni corrected *t*-tests have also been performed on *rendering condition* and the audio condition leads to relative errors that are significantly closer to 0% than the relative errors obtained with the other conditions (V,AV). Moreover, the relative errors ϵ_p are found to decrease significantly with increases of the *rendered distance* d_r . The interaction term that is found here reveals that ϵ_p decreases less with d_r for the audio condition than for the visual and audio-visual conditions, as is shown in Fig. 8.5(b). A perceptual distance compression phenomenon (overestimation for $d_r < 2.3$ m and underestimation for $d_r > 2.3$ m) is thus observed in the present experiment, and the observed compression is slightly, but significantly, less pronounced for the audio-only condition than for the visual and audio-visual conditions.

Post-session questionnaires

At the end of each session (A, V, and AV), subjects rated 15 statements on a 7-point Likert scale with two anchors (see Tab. 8.3). As differences between the different sessions are to be analyzed for each statement, any bias due to subjects has been removed using the following procedure: the rating $A_n^i(k)$ of the n^{th} subject for the k^{th} statements during session i ($i = A, V, AV$) has been transformed into $\underline{A}_n^i(k) = A_n^i(k) - M_n(k)$, with $M_n(k)$ the mean over the three sessions of the ratings of the n^{th} subject for the k^{th} statements.

Based on the unbiased ratings $\underline{A}_n^i(k)$, three scores related to *presence*, *spatial quality*, and *task difficulty* have been built. The P-score related to *presence* is the mean of the unbiased ratings $\underline{A}_n^i(k)$ across statements $k = [1, 2, 3, 4, 6, 9, 10]$. The SQ-score related to *spatial quality* is the mean of the unbiased ratings $\underline{A}_n^i(k)$ across statements $k = [5, 7, 8, 11, 12, 13]$. The TD-score related to *task difficulty* is the mean of the unbiased ratings $\underline{A}_n^i(k)$ across statements $k = [14, 15]$. The different statements and their contribution to each scores are recalled in Tab. 8.3.

Subject's scores related to *presence* (P-score), *spatial quality* (SQ-score), and *task difficulty* (TD-score) are shown in Fig. 8.6. A one-way ANOVA has been performed on each score with condition (A,V,AV) as a factor and with a 5% level. For the P-score, the one-way ANOVA shows that condition is highly significant with $F(2, 62) = 36.3$ and $p < 10^{-6}$. For the SQ-score, the one-way ANOVA shows that condition is highly significant with $F(2, 62) = 18.6$ and $p < 10^{-6}$. For the TD-score, the one-way ANOVA shows that condition is highly significant with $F(2, 62) = 20.7$ and $p < 10^{-6}$. For all scores, post-hoc tests (*t*-test with Bonferroni correction) reveal that the scores of the audio (A) condition are significantly lower than

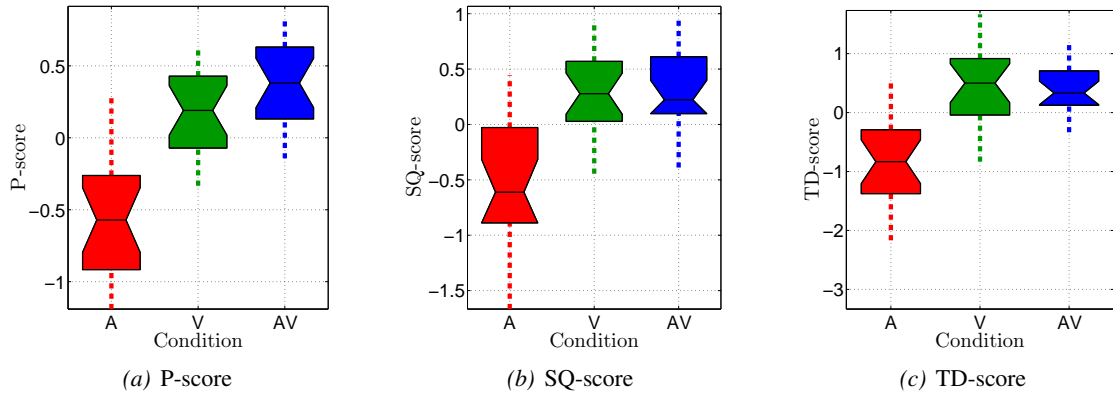


Figure 8.6: Subjects unbiased scores related to *presence* (P-score), to *spatial quality* (SQ-score) and to *task difficulty* (TD-score). See the caption of Fig. 8.4 for explanations regarding plot.

the scores of the others conditions (V and AV) and that the visual (V) and audio-visual (AV) conditions are not significantly different. Thus, *presence*, *spatial quality*, and *task difficulty* are rated significantly lower in the audio condition than in the visual and audio-visual. Moreover, *presence*, *spatial quality*, and *task difficulty* are rated equivalently for the visual and audio-visual rendering conditions.

Presentation order

To assess the effect of presentation order on any of the rendering conditions (A, V, AV) the following statistical analysis has been performed. Subject's relative errors ϵ_p have been separated into three groups, each divided into three subgroups, depending of the order of appearance of each condition: i) ϵ_p of the audio condition: audio first (A_1), audio second (A_2), and audio third (A_3); ii) ϵ_p of the visual condition: visual first (V_1), visual second (V_2), and visual third (V_3); iii) ϵ_p of the audio-visual condition: audio-visual first (AV_1), audio-visual second (AV_2), and audio-visual third (AV_3). For each of these groups (A_1, A_2, A_3 ; V_1, V_2, V_3 ; AV_1, AV_2, AV_3), a two-way ANOVA has been performed with the *order of appearance* (1, 2, 3) and the *rendered distance* d_r as factors.

Considering relative errors made during the audio condition, this analysis reveals that *rendered distance* d_r has a highly significant effect ($F(4, 327) = 97.1$ and $p < 10^{-6}$), that the *order of appearance* of the audio condition has a slightly significant effect ($F(2, 327) = 3.4$ and $p = 0.035$) and that there is no interaction effect between *order of appearance* and *rendered distance* ($F(8, 327) = 0.9$ and $p > 0.5$). Post-hoc tests (*t-test* with Bonferroni correction) show that during the audio condition subjects presented first with the audio condition made significantly less errors (for A_1 : $\langle \epsilon_p \rangle = -5.9\%$ and $SE(\epsilon_p) = 2\%$) than subjects presented with the audio condition in third position (for A_3 : $\langle \epsilon_p \rangle = -14\%$ and $SE(\epsilon_p) = 2.5\%$).

With respect to relative errors made during the visual condition, this analysis shows that *rendered distance* d_r has a highly significant effect ($F(4, 327) = 116$ and $p < 10^{-6}$), that the *order of appearance* of the visual condition does not have a significant effect ($F(2, 327) = 0.1$ and $p > 0.8$) and that there is a small degree of interaction effect between *order of appearance* and *rendered distance* ($F(8, 327) = 2.5$ and $p = 0.012$). This interaction effect is due to the rendered distance $d_r = 1.5$ m which is better estimated

by subjects presented with the visual condition in third position than the ones presented with the visual condition first during the visual condition.

Regarding relative errors made during the audio-visual condition, this analysis shows that *rendered distance* d_r has a highly significant effect ($F(4, 327) = 116$ and $p < 10^{-6}$), that the *order of appearance* of the audio-visual condition does not have a significant effect ($F(2, 327) = 1.6$ and $p > 0.2$) and that there is a small degree of interaction effects between *order of appearance* and *rendered distance* ($F(8, 327) = 2.8$ and $p = 0.004$). This interaction effect has the same explanation as the one observed in the visual condition.

A very small influence of the presentation order is thus observed in the present experiment. Only subjects presented with the audio condition in third position seemed to be affected and they made larger errors than subjects presented with the audio condition in the first position. However, this effect remains relatively small.

Learning effects

To assess the possibility of any learning effect that could have occurred during the 60 trials of the experiment, the following statistical analysis has been performed: relative errors ϵ_p made by the subjects have been divided in six groups of 10 successive trials (*i.e.* trials 1 to 10, trials 11 to 20, etc.). A one-way repeated-measures ANOVA has been performed on these six sets with the order of appearance of the 10-trials set as factor. This analysis reveals no significant differences between the relative errors made by the subjects among these different groups of trials with $F(5, 85) = 0.139$ and $p > 0.9$. Thus, no learning effect appears during the experiment.

8.2.4 Discussion

In this section, the hypotheses that have been made in Sec. 8.2.1 are discussed in the light of the experimental results given in Sec. 8.2.3.

Presence feeling and Visual distance underestimation

Hypothesis: because AV VR-systems (holophonic spatial audio and large immersive screens) provide a higher degree of *presence* they potentially lead to less visual distance underestimation than visual-only VR-systems (large immersive screens). Firstly, the potential influence of an auditory environment consistent with the rendered visual scene is analyzed. Secondly, the correlation between the *presence* feeling and visual distance underestimation is studied.

In experiments on visual distance estimation in virtual environments, a linear model is usually fit to the experimental data, as represented by Eq. (8.3) [Thompson et al. 2004, Klein et al. 2009, Grechkin et al. 2010]:

$$d_p = \alpha \times d_r + \beta \quad (8.3)$$

The dimensionless coefficient α , the *linear visual underestimation coefficient*, is thus the variable of interest in what follows.

In the experimental design depicted in Sec. 8.2.2 an auditory environment which is spatially coherent with the visual scene was proposed. However, the *presence* or *absence* of this coherent auditory environment has not been included as an independent variable. Results obtained in Sec. 8.2.3 are thus not directly applicable to assess the influence of a coherent auditory environment on the *presence feeling* and on the *linear visual underestimation coefficient* α . Nevertheless, the results obtained in the visual rendering condition can still be compared with similar results found in the literature.

In [Klein et al. 2009], a study on visual distance perception in VR-systems involving large immersive screens was carried out using a triangulation measurement protocol. The visual scene was similar to the one used here, but no coherent auditory environment was provided. In [Klein et al. 2009], a *linear visual underestimation factor* $\alpha = 31.4\%$ was found with $\beta = 0.97$ m and the linear model of Eq. (8.3) explains $R^2 = 93.5\%$ of the variance present in the experimental data. By fitting a linear model in the form of Eq. (8.3) to the experimental results of Sec. 8.2.3 in the visual rendering condition, a *linear visual underestimation coefficient* $\alpha = 28.2 \pm 5.6\%$ is found with $\beta = 1.43 \pm 0.18$ m. The linear model explains $R^2 = 89.2\%$ of the data.

The results obtained here are thus in very good agreement with the results of [Klein et al. 2009]. The factor β is however found to be larger in the present study (see Sec. 8.2.4 for more comments on this point). It seems that the presence of a coherent auditory environment does not have a large effect (if any) on the *linear visual underestimation factor*. An experimental design which includes explicitly the *presence* or *absence* of the coherent auditory environment as an independent variable is however needed to conclude more firmly on this point.

The main problem of the feeling of *presence* is that it is subject-dependent. This is illustrated in Fig. 8.6(a) where mean P-scores in the visual and audio-visual rendering conditions are not found to be significantly different. Actually, for some of the subjects, *presence* was higher in the audio-visual rendering condition than in the visual rendering condition. For others, the opposite was true. This is potentially a consequence of the chosen audio stimulus (low pass filtered white noise, see 8.2.2) which is perceived as unpleasant by some subjects and thus decreased their feeling of *presence*.

It is thus more appropriate to study P-scores subject by subject. For each subject, the *presence* variation Δ_P induced by the addition of the spatialized audio stimuli is calculated as the difference between the P-score of that subject in the audio-visual condition and the P-score of that subject in the visual condition. Similarly, the *linear visual underestimation factor variation* Δ_α induced by the addition of the spatialized audio stimuli is calculated as the difference between the *linear visual underestimation factor* of that subject in the audio-visual condition and the *linear visual underestimation factor* of that subject in the visual condition. Δ_α is represented as a function of Δ_P for all the subjects in Fig. 8.7. It is noticed that, as discussed previously, Δ_P is not necessarily positive.

Taking into account data of all subjects, Δ_α and Δ_P are not found to be correlated (correlation coefficient of $\Gamma = 0.27$ and $p = 0.36$). However, as can be seen in Fig. 8.7, four points are far away from the remaining points. Taking the data of all subjects except these four in account, Δ_α and Δ_P are found to be significantly correlated with a correlation coefficient of $\Gamma = 0.80$ and $p = 0.001$. Thus for more than 75% of the subjects (13 out of 17), Δ_α and Δ_P are significantly correlated. This result supports the idea that a higher degree of *presence* can lead to less visual distance underestimation and illustrates again inter-subject variability. Nevertheless, the order of magnitude of variations of *linear visual underestimation factor* induced by the

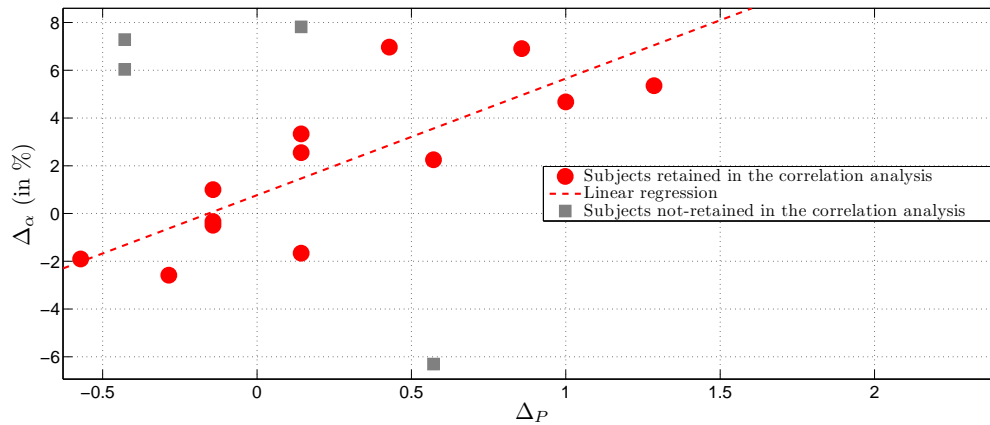


Figure 8.7: *Linear visual underestimation factor variation Δ_α induced by the addition of the spatialized audio stimuli as a function of the feeling of *presence* variation Δ_P induced by the addition of the spatialized audio stimuli for each subject.*

feeling of *presence* variation is not large enough to explain the overall distance underestimations reported here.

Perceived distance in *audio* VR-systems based on *holophonic* sound rendering

The objective of this section is to discuss how distances to static sound sources simulated by *holophonic* spatial sound renderings, and more specifically by WFS, are perceived by moving subjects in comparison to distances to real static sound sources.

The advantage of *holophonic* spatial sound rendering technologies is that they can synthesize the acoustical wave field corresponding to a given virtual audio source, including its curvature. For static listeners, this curvature induces binaural differences which are useful for estimating distances to near sources (< 1 m) [Kim et al. 2001, Wittek et al. 2004]. For moving listeners, this curvature allows *audio* motion parallax to be naturally rendered. This is interesting in the present context because *audio* motion parallax has been shown to provide several useful cues for distance estimation [Speigle & Loomis 1993].

Nevertheless, *holophonic* spatial sound rendering technologies, such as WFS, suffer *in practice* from a number of drawbacks which limit the perceptual quality of the synthesized wave field [Corteel 2004, Wittek et al. 2004]. The aliasing frequency f_{al} is the upper limit above which the wave field is neither spatially nor spectrally correctly reconstructed (see Sec. 8.2.4). Among other parameters, f_{al} depends on the listener's position in the rendering area. Moreover, within the listening area, the distance intensity decay of a virtual source rendered using WFS fails to meet the $1/r$ -law and follows a $1/\sqrt{r}$ -law instead [Corteel 2004, Wittek et al. 2004]. In practice, listeners moving in the *audio virtual* world rendered using WFS do not therefore receive exactly the same acoustical cues that they would have received in the *real* world.

Distance perception to near sources rendered using WFS (0.25 m to 1.9 m) in the case of fixed listeners has been studied by [Wittek et al. 2004]. It has been concluded that the curvature of the wave front of WFS virtual sources was perceptually not sufficiently well reconstructed for distance perception in this case. In [Corteel 2004], it has however been shown that WFS can provide moving listeners a good perspective

impression. In [Rébillat et al. 2008], it was furthermore shown that WFS can convey a certain sense of distance to moving listeners when only wave front curvature is available. Results from these studies [Corteel 2004, Rébillat et al. 2008] were not comparable to real world perception because no equivalent real-world studies have been performed.

The present study assesses the ability of a moving listener to judge the egocentric *audio* distances to static virtual objects synthesized using WFS by using triangulation. A compressive model in the form of $d_p = k(d_r)^a$ is a good candidate for the psychophysical function that relates estimates of perceived distance to physical source distance [Zahorik et al. 2005]. The experimental protocols [Loomis et al. 1998] (verbal report, perceptually directed action, ...) and the listening conditions [Speigle & Loomis 1993] (static or moving) have been shown to have very little influence on the values of k and a . Results obtained in Fig. 8.5(a) for the *audio* condition can thus be compared directly with real world studies involving static listeners. In this context, *audio* distance perception has been widely studied and a review of the relevant literature is presented in [Zahorik et al. 2005]. The histograms of the values for k and a obtained among 84 data sets are presented in [Zahorik et al. 2005, Fig. 1]. Values that appeared most frequently in the histogram are between 0.3 and 0.4 for a and between 0.7 m and 1 m for k .

By fitting a compressive model to the perceived audio distances collected in the present study, values of $a = 0.34 \pm 0.085$ and $k = 1.69 \pm 0.31$ m are found, with $R^2 = 98\%$ of the variance observed in the experimental data explained by the compressive model. The model $d_p = k(d_r)^a$ thus fits very well to the experimental data. Moreover, the value found for a is highly consistent with values resulting from real-world experiments in the literature. The value found for k is slightly larger than the most frequently appearing value of [Zahorik et al. 2005, Fig. 1] but is still consistent with some values coming from real world studies (see Sec. 8.2.4 for more comments on this point).

As shown in Tab. 8.4, results for standard deviation σ found in the present experiment under the audio condition and presented in Fig. 8.5(a) are also in good agreement with standard deviations found in real-world studies of audio distance perception involving perceptually directed action protocols.

[Speigle & Loomis 1993, Fig. 2, dynamic 2 m, 90°]	d_r	2 m	-	4 m	-	6 m
	σ	$\simeq 0.5$ m	-	$\simeq 0.7$ m	-	$\simeq 0.7$ m
[Loomis et al. 1998, Fig. 1, Audition, XP 1, gymnasts]	d_r	3 m	-	-	-	8 m
	σ	$\simeq 0.5$ m	-	-	-	$\simeq 0.5$ m
Present study	d_r	1.5 m	2 m	2.7 m	3.5 m	5 m
	σ	0.45 m	0.50 m	0.53 m	0.66 m	0.79 m

Table 8.4: Comparison of standard deviations obtained in real and virtual environments.

It can thus be concluded that despite its inherent drawbacks, WFS is able to synthesise sound-fields which are perceptually meaningful in terms of distance for *moving subjects* and for static *virtual* sources placed in the *action space*. Moreover, the *virtual* WFS sound field provides a perception of distance that is as compressed as is the *real* world perception of distance.

Audio-visual distance perception by moving subjects in virtual environment

In Sec. 8.2.3, results are presented for rendered distances between 1.5 m and 5 m estimated using triangulation after an exploration phase where subjects were allowed to move in a restricted area. These results

show that perceived distances are systematically overestimated for rendered distances closer than 2.3 m and underestimated for distances farther than 2.3 m. Interestingly, results also show that all the rendered distances are perceived consistently across the different available conditions. These results are in good agreement with the results of [Côté et al. 2011] which suggested that static subjects perceived AV distances similarly to visual distances in virtual environments.

The fact that subjects take benefit from AV dynamic distances cues does not lead to results that are qualitatively different from the static case in terms of estimated distance. Nevertheless, it is important to notice that, as shown in Sec. 8.2.3, all of the subjects spontaneously walked during the exploration phase and that the exploration durations t_{XP} were different among the different modalities, with $t_{XP}(A) > t_{XP}(AV) > t_{XP}(V)$. Subjects thus attempted to gain information from the AV dynamic cues and seemed to proceed differently depending on the available modality. Moreover, they walked almost twice as far in the direction perpendicular to the virtual object than in the direction parallel to the virtual object during the exploration phase. This result strongly hints at the use of motion parallax (changes in angular direction of a point source occasioned by the subject's translation, see Sec. 8.2.2) rather than acoustic τ (time-to-collision for a constant velocity between subject and source, see Sec. 8.2.2) for distance estimation.

It can thus be hypothesized that subjects are trying to use all the available distance cues before giving their distance estimation. The applied strategy and time needed for the processing of the different cues furthermore appear to be modality-dependent.

Influence of the physical presence of the AV RV-system on the perceived virtual space

In this section, the fact that the AV VR-system used in the present experiment is not fully transparent, and the possibility of a resulting spatial link existing between the real and virtual worlds are discussed.

The WFS technology used here to render the spatialized sound is not perfect and conflictual audio spatial cues are unavoidably presented. The main drawback of WFS comes from the aliasing frequency f_{al} , which is the upper limit above which the wave-field is neither spatially nor spectrally correctly reconstructed. In the setup used here, $f_{al} = 1070 \pm 70$ Hz and the chosen audio stimulus contains frequencies up to 4 kHz. To localize the virtual object in azimuth when travelling in the exploration area, humans use several different acoustic cues [Blauert 1999]. In the low frequency range (*i.e.* below $\simeq 1.5$ kHz) interaural phase differences (IPD) are the dominant cues. In the high frequency range, interaural envelope differences (IED) and interaural loudness differences (ILD) become dominant. The AV RV-setup used here thus allows to render IPD cues corresponding to the position of the virtual object, whereas IED and ILD cues specify the position of the physical setup. The 4 kHz low-pass filtering described in Sec. 8.2.2 has thus been introduced in order to limit the amount of energy present above the aliasing frequency, and thus to limit possible inconsistencies between low- and high- frequency audio azimuthal cues. Nevertheless, it has been shown in [Wightman & Kistler 1992] that IPD has a dominant role in sound localization for broadband stimuli. Moreover, in [Sanson et al. 2008] it was demonstrated that sound fields reconstructed by WFS are sufficiently consistent to allow for accurate localization, even when frequencies above 1.5 kHz were present. In [Cortel et al. 2007], it was shown that even if not *exact*, azimuthal cues above the aliasing frequency f_{al} were generally consistent with azimuthal cues below f_{al} when using MAPs. For these reasons, the possible inconsistencies between low- and high- frequency audio azimuthal cues should not prevent subjects from

correctly localizing the virtual objects in azimuth.

Another potential conflicting audio spatial cue concerns reverberation. As specified in Sec. 8.2.2, no room effect was synthesized in order to recreate acoustical conditions that were as close as possible to open free-field conditions. Nevertheless, even though the experimental room had been acoustically treated, there were still traces of a room effect, with a mid-frequency mean reverberation time T_{60} (500 Hz to 1 kHz) of 0.45 s. The ratio of the energies of the direct and reverberated parts of the sound, which is an audio distance cue [Bronkhorst & Houtgast 1999], specifies to the subject a distance corresponding to the physical setup rather than of the virtual object. The acoustic reverberation distance cue is thus in conflict with other distance cues.

The technology used to provide the 3D visual rendering is also not perfect either. In tracked passive stereoscopy, two images of a given scene taken from two slightly different points of view are projected on a planar screen. To estimate the distance of the virtual visual object presented this way, subjects use two binocular cues. Focus cues (accommodation and blur in the retinal image) specify the distance at which the screen, instead of the virtual object, is seen. Vergence cues corresponds to the distance at which the optical axes of the two eyes cross one another, *i.e.* the virtual object. In a 3D visual rendering setup based on large immersive screens, focus cues are then almost always in conflict with the vergence cues [Howarth 2011]. This *vergence-accomodation* conflict can cause visual fatigue [Hoffman et al. 2008] and can affect depth perception if the distances specified by the two cues are too different [Watt et al. 2005].

The AV VR-system used here is thus not perfect and provides subjects with some conflictual AV spatial cues. As has been explained above, these conflicts should not prevent the subjects from correctly localizing the virtual objects, but they can decrease the *transparency* of the AV VR-system by reminding its physical presence to the subjects.

In the results presented in Fig. 8.5, the point at which rendered distance equals perceived compressed *audio*, *visual*, or *audio-visual* distances is located at 2.3 m. This distance is the distance between the subject in the starting position and the right LaMAP (the flat array of loudspeakers used as a front projection screen, see Sec. 8.2.2). When a virtual object is rendered at a distance of 2.3 m, its image and sound position are thus *physically* and *perceptually* close to the position of the LaMAP. We propose to say that “*the virtual world is anchored to the real world*” at this point.

This proposed *anchor* effect could also be noticeable in Sec. 8.2.3, Fig. 8.4(b), where it has been shown that significantly less time was spent exploring the virtual environment with far rendered distances (3.5 m and 5 m) than for close distances (1.5 m, 2 m and 2.7 m). Shadows projected on the ground floor were visible for $d_r = 3.5$ m and $d_r = 5$ m whereas they were not for $d_r = 1.5$ m, $d_r = 2$ m and $d_r = 2.7$ m. This could explain the larger amount of time needed to estimate distances of the closer virtual objects. The presence or absence of the projected shadow is a direct consequence of the physical presence of the screen. It can also contribute to the *anchor* phenomenon by reminding the physical presence of the AV VR-system to the subjects.

This *anchor* phenomenon can be considered in light of the so-called *specific distance tendency*, defined as the distance at which an subject perceives an object in the absence of any distance information [Gogel 1969]. As stated in [Zahorik 2002], “*an estimate of this distance tendency may be computed by determining the point at which perceived distance is equivalent to physical distance*”. So, in the present study the AV *specific distance tendency* is displaced to the distance at which the AV physical setup is positioned and thus

to the distance where the virtual world is anchored to the real world. This is a possible explanation for the values of β and k found in Secs. 8.2.4 and 8.2.4 that are greater than values often reported in the literature.

It is thus hypothesized that this *anchor* phenomenon scales the audio and visual virtual environments according to the position of the AV VR-system in the real environment. This can be interpreted as a consequence of the limited *transparency* of the AV VR-system caused by conflicting AV distance cues available to the subjects.

8.2.5 Conclusion

In this paper a study of *audio*, *visual*, and *audio-visual* egocentric distance perception by moving subjects in *virtual* environments is presented. Audio-visual rendering was provided by tracked passive visual stereoscopy and acoustic wave field synthesis (WFS). For each rendering condition, the estimation of perceived distances was based on a perceptually directed action using the method of indirect blind-walking.

Distances perceived in the virtual environment were systematically overestimated for rendered distances closer than 2.3 m and underestimated for distances farther than 2.3 m. Interestingly, all rendered distances were similarly perceived across the different available conditions.

Distance underestimations observed in visual virtual environments appeared to be linked to the feeling of *presence*: the more *present* the subject felt, the less *underestimated* the distance. In addition, the audio-visual rendering system was not found to be fully transparent and it is hypothesized that the rendering system induced an *anchor effect* between the virtual and real environments.

Acknowledgments

The authors thank all the volunteers who took part in the experiment. Thomas Chartier and Philippe Cuvillier, currently students at the École Polytechnique (France), are also thanked for their help in the design and preliminary tests. Additionally, special thanks is given to Matthieu Courgeon for the time he spent on the visual rendering. Finally, the authors would like to thank Antonio Trujillo-Ortiz [Trujillo-Ortiz et al. 2004] for providing a reliable Matlab version of the repeated-measures two-way analysis of variance test.

8.3 Discussion

Cette discussion générale propose d'abord de comparer les résultats audio-visuels obtenus dans la section 8.2 aux prédictions effectuées par la théorie statistique de l'intégration audio-visuelle présentée à la section 7.3.3. Ensuite, les limitations du protocole d'estimation de la distance perçue choisi ici, à savoir la procédure de triangulation sont étudiées. Cette discussion s'achève sur une comparaison des résultats obtenus lors des deux expériences de perception de la distance présentées dans les sections 7.4.4 et 8.2.

8.3.1 Retour sur l'intégration audio-visuelle en distance

Cette partie de la discussion s'attache à mieux comprendre les phénomènes d'interaction audio-visuelle en distance intervenant dans l'expérience décrite à la section 8.2. Dans un premier temps, la variabilité (et donc la fiabilité perceptive) des estimations de distance réalisées en utilisant les modalités auditive et visuelle est analysée. Ensuite, les résultats obtenus pour l'estimation de distance audio-visuelle sont comparés à des résultats prédits par un modèle statistique d'intégration audio-visuelle basé sur les performances unimodales (voir section 7.3.3).

Variabilité des estimations de distance auditive et visuelle

Il est bien établi que les jugements de localisation en azimuth basés sur la modalité visuelle sont plus fiables (c'est à dire moins variables) que ceux réalisés en utilisant la modalité auditive (voir les sections 7.2.1 et 7.2.2 par exemple). Cela est aussi vrai pour les jugements de distances proches ($i.e < 1$ m) [Bowen 2010, Agganis et al. 2010], mais ne semble pas être le cas pour les distances plus importantes [Loomis et al. 1998].

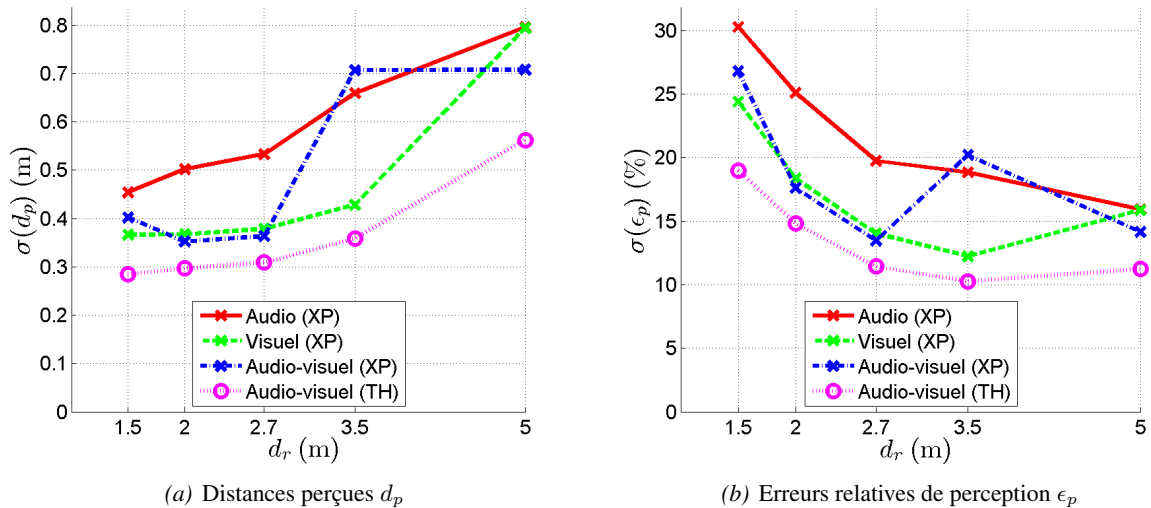


FIG. 8.8 – Écarts-types expérimentaux (XP) obtenus à partir des distances perçues d_p et des erreurs relatives de perception ϵ_p faites en utilisant les modalités audio, visuelle et audio-visuelle en fonction de la distance de rendu d_r . Les écarts-types prédits par une théorie bayésienne de l'intégration audio-visuelle sont aussi tracés (TH).

Les données collectées à la section 8.2.3 peuvent être analysées sous cet angle. Sur la figure 8.8, les écarts-types associés aux distances perçues d_p et aux erreurs relatives de perception ϵ_p sont tracés, entre autres, pour les modalités auditive et visuelle. Ces figures montrent que les écarts-types liés à la modalité visuelle sont toujours inférieurs ou égaux à ceux liés à la modalité auditive. Cependant, elles montrent aussi que ces deux écarts-types sont globalement du même ordre de grandeur. Ainsi, ces données suggèrent que la modalité visuelle est seulement légèrement plus fiable que la modalité auditive pour l'estimation de distances comprises entre 1 m et 5 m avec le protocole de triangulation utilisé ici. Cette différence de fiabilité s'estompe cependant lorsque la distance des objets à localiser augmente.

Comparaison avec les résultats prédits par une théorie de l'intégration audio-visuelle

Il semble que le cerveau utilise les données issues de chacun des sens et les pondère selon leurs fiabilités respectives selon un modèle de maximum de vraisemblance (voir section 7.3.3) [Perrott 1993, Battaglia et al. 2003, Alais & Burr 2004, Nguyen 2008, Godfroy et al. 2011]. Comme il a été vu précédemment, l'être humain est capable de déterminer une estimation m_V de la distance à laquelle se situe un objet visuel avec une variance σ_V , et une estimation m_A de la distance à laquelle se situe un objet auditif avec une variance σ_A . Selon la théorie mentionnée précédemment, la distance à laquelle se situe un objet audio-visuel est alors estimée autour d'une position moyenne m_{AV} et avec un écart-type σ_{AV} données par :

$$m_{AV} = \frac{\sigma_V^2}{\sigma_V^2 + \sigma_A^2} m_A + \frac{\sigma_A^2}{\sigma_V^2 + \sigma_A^2} m_V \quad \text{et} \quad \sigma_{AV}^2 = \frac{\sigma_A^2 \sigma_V^2}{\sigma_A^2 + \sigma_V^2} \leq \min(\sigma_A, \sigma_V) \quad (8.4)$$

Comme le montre la figure 8.8, $\sigma_A \simeq \sigma_V$ dans le cas présent. Ainsi, selon cette théorie, l'audition et la vision seraient utilisés conjointement pour donner des résultats plus fiables en contexte audio-visuel qu'en contexte unimodal.

Sur la figure 8.8 sont tracés les écarts-types observés expérimentalement en contexte audio-visuel et ceux prédits par la théorie à l'aide de l'Eq. (8.4) pour les distances perçues d_p et pour les erreurs relatives de perception de la distance ϵ_p . Les prédictions théoriques semblent ici globalement beaucoup trop optimistes par rapport aux résultats expérimentaux. Sur la figure 8.9 sont tracées les moyennes observées expérimentalement en contexte audio-visuel et celles prédites par la théorie à l'aide de l'Eq. (8.4) pour les distances perçues d_p et pour les erreurs relatives de perception de la distance ϵ_p . Les prédictions théoriques semblent ici en bon accord avec les données expérimentales.

Bilan

Fiabilité (écarts-types) : il est possible de conclure que la théorie de l'intégration audio-visuelle exposée précédemment sous-évalue les écarts-types par rapport aux résultats expérimentaux obtenus ici. Une des raisons pouvant expliquer cet optimisme est que le protocole de report utilisé ici (la triangulation) fait intervenir une réponse *motrice* dans la procédure d'estimation de la distance. La procédure de triangulation induit donc inévitablement une composante *motrice* dans la variabilité en plus de la composante *perceptive* que l'on cherche à estimer ici. Si la composante *motrice* est d'un ordre de grandeur comparable à celui de la composante *perceptive*, alors il paraît normal que la théorie de l'interaction audio-visuelle utilisée ici, qui ne modélise pas cette variabilité *motrice*, ne fasse pas de bonnes prédictions. Pour pouvoir obtenir des

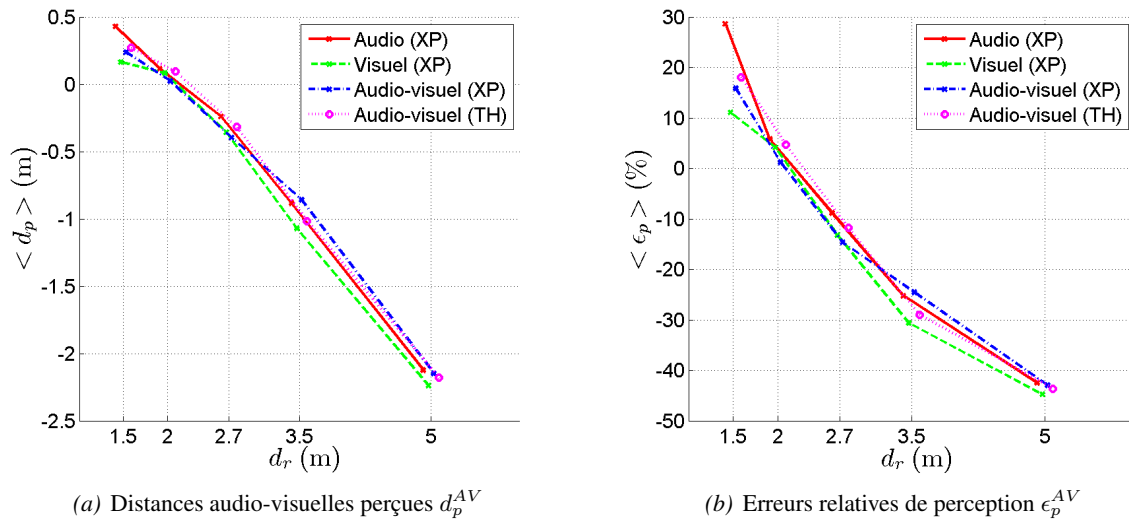


FIG. 8.9 – Comparaison des moyennes calculées sur les données de perception audio-visuelles obtenues expérimentalement et des moyennes prédites par la théorie de l'intégration audio-visuelle.

résultats fiables sur ce point, il est donc nécessaire d'une part de savoir précisément avec quelle variabilité les sujets sont capables d'effectuer un pas dans une direction déterminée par un objet audio ou visuel et d'autre part d'être capable d'intégrer cette variabilité dans le modèle d'intégration. Le premier de ces deux points sera étudié plus en détail à la section 8.3.2.

Valeurs moyennes : la théorie semble ici en bon accord avec les données expérimentales. Il faut cependant nuancer assez fortement ce résultat. En effet les analyses de la section 8.2.3 ne révèlent aucune différence significative entre les distances perçues d_p auditives, visuelles et audio-visuelles. Et la théorie d'intégration audio-visuelle utilisée ici se contente de calculer un barycentre entre les estimations moyennes des distances perçues auditives et visuelle. Dans ce contexte il est donc assez prévisible d'obtenir un bon accord entre théorie et expérience.

8.3.2 Limites de la procédure de triangulation (étude préliminaire)

La procédure d'estimation de la distance perçue utilisée à la section 8.2 est une procédure dite *d'action perceptuellement dirigée*, en opposition aux procédures de *report verbal* et *d'action temporelle imaginée* (voir section 8.2.1). Ce type de procédure a été choisi car, en comparaison des procédures de *report verbal* elle permet de limiter la variabilité des réponses des sujets. Par rapport aux procédures *d'action temporelle imaginée*, elle permet aussi d'éviter un biais systématique pouvant apparaître lors de la conversion d'informations temporelles en informations de distance. Cependant, comme cela a été mentionné dans la section 8.3.1, cette procédure n'est pas parfaite. En effet, elle se base sur une réponse motrice des sujets à un stimulus perceptif donné : ils doivent *effectuer un pas en avant* dans la direction où ils pensent que se situe l'objet qui leur a été présenté (voir section 8.2.2). Cette réponse motrice présentant inévitablement une certaine variabilité, l'expérience préliminaire présentée ci-dessous vise à quantifier cette variabilité *motrice* pour pouvoir la comparer à la variabilité *perceptive* recherchée à la section 8.3.1.

Protocole expérimental

Les figures 8.8(a) suggèrent que la variabilité calculée sur les distances perçues est *linéairement* croissante avec la distance rendue d_r . Cela semble donc indiquer que la variabilité *angulaire* associée à un pas fait dans une direction donnée est constante. L'hypothèse testée ici est alors que lorsque les sujets effectuent un pas en avant dans une direction donnée, il existe une incertitude angulaire sur la direction dans laquelle ils se dirigent qui soit indépendante de la distance de l'objet et indépendante de la modalité de rendu. Cette incertitude serait purement *motrice*.

Le protocole expérimental utilisé est quasiment le même que celui présenté à la section 8.2.2. Les seules différences sont que les sujets ne doivent plus fermer les yeux lors de la procédure de triangulation et que les objets virtuels présentés sont toujours actifs au cours d'un essai. De plus, seules les modalités auditive et visuelle ont été testées ici. Un total de 5 sujets a participé à cette expérience (uniquement des hommes déjà volontaires pour l'expérience décrite à la section 8.2). Les distances testées sont les mêmes que celles testées dans la section 8.2, à savoir 1.5 m, 2 m, 3.5 m et 5 m. Chaque distance a été testée 3 fois et les ordres des sessions *audition* et *vision* sont alternés pour chaque sujet.

Les variables indépendantes sont ici la distance rendue d_r et la modalité de rendu. La variable dépendante est la partie variable de l'erreur angulaire $\underline{\Delta}_\alpha$ effectuée par chaque sujet lorsqu'il effectue un pas dans la direction de la source. Pour chaque essai, la partie variable de l'erreur angulaire $\underline{\Delta}_\alpha$ est définie comme l'erreur angulaire Δ_α effectuée par le sujet à cet essai diminuée de la moyenne des erreurs angulaires $\langle \Delta_\alpha \rangle$ réalisées à cette distance et dans cette modalité par tous les sujets : $\underline{\Delta}_\alpha = \Delta_\alpha - \langle \Delta_\alpha \rangle$.

Résultats

La partie variable de l'erreur angulaire effectuée par les sujets lorsqu'ils effectuent un pas dans la direction de la source est analysée à l'aide d'une ANOVA à mesures répétées avec la distance de rendu d_r et la modalité de rendu comme facteurs. Les résultats indiquent que la distance de rendu d_r n'affecte pas significativement la partie variable de l'erreur angulaire ($F(4, 1) = 0.44$ et $p = 0.78$), que la modalité non plus ($F(1, 4) = 3.77$ et $p = 0.12$), et enfin qu'il n'existe pas d'interaction significative entre la distance de rendu et la modalité ($F(4, 4) < 10^{-4}$ et $p \simeq 1$). Lorsqu'il leur est demandé d'effectuer un pas dans la direction d'un objet audio ou visuel, les sujets sont donc capables de se diriger dans la direction de cet objet à $\pm 4^\circ$ (écart-type de la partie variable de l'erreur angulaire calculé sur tous les essais) indépendamment de la distance et de la modalité de restitution de l'objet.

Les résultats obtenus ici méritent cependant d'être interprétés avec prudence. En effet, relativement peu de sujets ont participé à cette expérience et le fait d'avoir utilisé des objets visuels et auditifs virtuels peut avoir eu de l'influence sur les réponses des sujets. Il serait donc intéressant de réaliser une expérience similaire mais avec des objets réels et un plus grand nombre de sujets pour confirmer l'ordre de grandeur obtenu ici.

Comparaison des variabilités *motrices* et *perceptives*

À la section 8.3.1 il a été suggéré que si, dans la variabilité *perceptuo-motrice* observée à la section 8.2.3, la composante *motrice* est du même ordre de grandeur que la composante *perceptive*, alors la théorie

de l'interaction audio-visuelle (qui ne modélise pas cette variabilité *motrice*) ne permet pas de faire de bonnes prédictions. Sur la figure 8.10 sont tracés les écarts-types des erreurs angulaires signées pour chaque modalité et pour deux expériences subjectives différentes. Les deux expériences subjectives comparées ici sont celle décrite à la section 8.2 évaluant la variabilité angulaire *perceptuo-motrice* et celle présentée ici évaluant a priori uniquement la variabilité angulaire *motrice*.

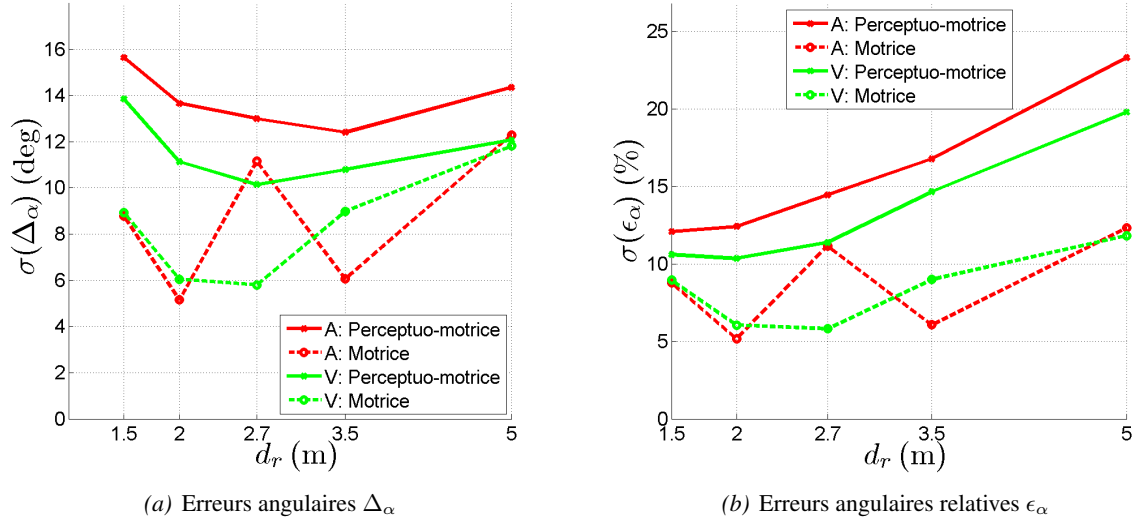


FIG. 8.10 – Écarts-types calculés sur les erreurs angulaires Δ_α dans les cas où la procédure expérimentale fait intervenir une réponse *motrice* et une réponse *perceptuo-motrice* auditive ou visuelle.

Les résultats de la figure 8.10 montrent que pour chacune des modalités, les composantes *motrices* de la variabilité sont inférieures aux composantes *perceptuo-motrices* de la variabilité observée¹, ce qui semble assez logique. D'autre part, cette figure illustre bien le fait que dans l'expérience réalisée, les composantes *motrices* de la variabilité angulaire (voir figure 8.10(a)) sont indépendantes de la modalité et à peu près indépendantes de la distance, ce qui n'est pas le cas des composantes *perceptuo-motrices* de la variabilité angulaire qui, en outre, dépendent de la modalité. Enfin, vu l'ordre de grandeur de la composante *motrice* par rapport à l'ordre de grandeur de la composante *perceptuo-motrices*, les écarts entre théorie et expérience observés sur la figure 8.8, ne sont vraisemblablement pas le fruit d'une déficience de la théorie de l'intégration audio-visuelle, mais plutôt celui de la variabilité de la méthode de report qui n'est pas prise en compte dans le modèle utilisé. Il serait donc intéressant d'essayer d'inclure cette part de variabilité (indépendante de la distance de rendu d_r et de la modalité de rendu) dans le modèle théorique d'interaction audio-visuelle pour obtenir de meilleures prédictions.

8.3.3 Comparaison des deux expériences d'estimation de la distance audio

Le dernier point qu'il est intéressant de discuter dans ce chapitre est la comparaison des résultats obtenus lors des deux expériences subjectives d'estimation de la distance auditive présentées dans les sections 7.4.4

1. La brusque variation observée pour $d_r = 2.7$ m est certainement due au faible nombre de sujets ayant participé à l'expérience.

et 8.2. Sur la figure 8.11 sont tracés les moyennes et écarts-types obtenus lors de l'expérience présentée à la section 7.4.4 (restitution de l'effet de parallaxe), ceux obtenus pour la modalité auditive à la section 8.2, et la courbe de compression moyenne issue de la synthèse bibliographique de Zahorik et al. [2005] (modèle du type $d_p = kd_r^a$ avec $k = 1.32$ et $a = 0.54$).

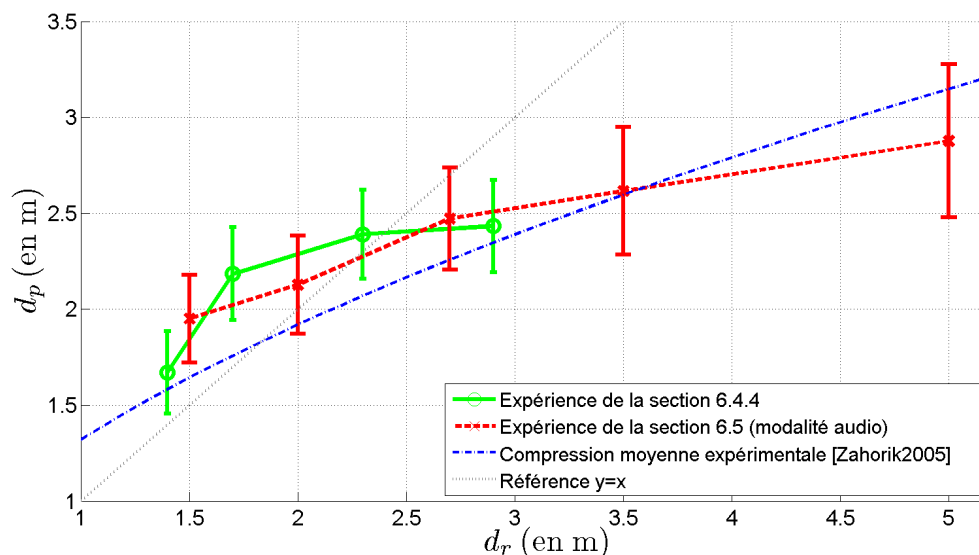


FIG. 8.11 – Comparaison des résultats obtenus durant les expériences subjectives d'estimation de la distance décrites à la section 7.4.4 et à la section 8.2. Les barres verticales indiquent les écarts-types.

L'étude de la figure 8.11 révèle que les distances perçues par les sujets lors des deux expériences sont globalement en bon accord entre elles et avec la littérature, et ce malgré des différences importantes existant entre les deux protocoles expérimentaux :

1. La première de ces différences réside dans le fait que ces expériences ont été réalisées sur deux versions successives du SMART-I² : la version ne possédant qu'un seul panneau, une égalisation unique pour tous les excitateurs, et située dans une petite salle pour l'expérience la section 7.4.4 (voir figure 7.11), et la version possédant les deux panneaux, une égalisation par excitateur, et située dans une grande salle pour l'expérience de la section 8.2 (voir figure 2.6). Le fait que les estimations perceptives de distances soient cohérentes entre les deux cas semble donc indiquer que l'espace virtuel auditif restitué par le SMART-I² est relativement indépendant du lieu de restitution.
2. La deuxième de ces différences se situe dans la méthodologie utilisée pour réaliser les estimations de distance. Dans la première expérience (section 7.4.4), les sujets rapportaient la distance auditivement perçue par l'intermédiaire d'un report dans la modalité visuelle. La seconde expérience utilise quand à elle une méthode de report basée sur une réponse motrice des sujets. Il semble donc que ces deux méthodes de report donnent des résultats comparables, que ce soit en terme de précision ou bien en terme de variabilité. Il serait cependant nécessaire d'évaluer ce point de façon plus méthodique pour pouvoir conclure plus fermement.
3. La dernière différence se situe dans le nombre d'indices acoustiques disponibles pour estimer la distance des objets sonores virtuels. En effet, l'expérience de la section 8.2 permet aux sujets de

se baser sur l'intensité acoustique tandis que l'expérience de la section 7.4.4 ne le leur permet pas. Cela explique sans doute le fait que sur la figure 8.11 la compression de l'espace auditif semble être plus accentuée pour l'expérience de la section 7.4.4 que pour celle de la section 8.2. Comme pour le point précédent, une validation subjective rigoureuse est indispensable pour pouvoir confirmer cette hypothèse.

8.4 Conclusion

Ce chapitre a cherché à comprendre comment la *perception* de l'environnement audio-visuel synthétisé par le SMART-I² était liée à la *connaissance* qu'ont les sujets du dispositif physique. Une expérience explorant l'influence des limitations de la *transparence* du SMART-I² sur la perception de la distance et sur la sensation de *présence* a donc été réalisée. La spécificité de cette expérience, en regard de la littérature, est de proposer aux sujets une phase d'exploration durant laquelle ils sont susceptibles d'acquérir un maximum d'informations concernant le monde virtuel qui leur est proposé.

Du point de vue *perceptif* cette expérience a montré que, de façon analogue à l'expérience présentée à la section 7.4.4 et à la littérature, les distances perçues dans l'environnement virtuel étaient systématiquement sous-estimées pour des distances de rendu inférieures à 2.3 m et sur-estimées pour des distances de rendu supérieures à 2.3 m. Étonnamment, les distances sont perçues de façon similaire quelle que soit la combinaison de modalités impliquée (audio, visuel ou audio-visuel). Il semble de plus que le protocole de report utilisé (la triangulation) induise une variabilité *motrice* dans les estimations de distances effectuées par les sujets. Les résultats multi-modaux ne peuvent alors pas être prédits par une théorie statistique simple.

Du point de vue d'un éventuel lien entre *perception* et *cognition*, on peut tirer deux conclusions :

- La sous-estimation des distances est liée à la sensation de *présence* : plus le sujet se sent *présent* dans le monde virtuel, moins il sous-estime les distances dans le monde virtuel. Or, pour que la sensation de *présence* soit forte, le sujet doit impérativement *croire* à ce qui lui est présenté et oublier ce qu'il *sait* (à savoir que le monde qu'il voit et entend est *virtuel*). Les résultats obtenus suggèrent donc que plus le sujet arrive à lâcher prise sur sa *cognition*, plus sa *perception* du monde virtuel s'étend spatialement.
- Le système de restitution audio-visuelle n'est pas complètement *transparent* d'un point de vue perceptif : il est alors supposé que de par sa présence physique, le SMART-I² *ancre* le monde virtuel dans le monde réel. Cela suppose alors que même si le sujet arrive à complètement lâcher prise sur sa *cognition*, les imperfections du rendu viendront toujours lui rappeler que son expérience est virtuelle.

Le protocole expérimental développé pour cette expérience autorise de nombreuses variations permettant d'aborder de nouvelles perspectives. La première perspective proposée ici concerne la comparaison de différents types de rendu sonores spatialisés. L'expérience décrite à la section 8.2 a été menée en utilisant la "Wave Field Synthesis" comme système de rendu sonore spatialisé. Il est cependant possible et assez aisé de réaliser la même expérience en incluant le système de rendu sonore spatialisé comme variable indépendante (voir la section 2.4.2 et la table 2.2 pour un aperçu des différentes technologies de rendu sonore spatialisé disponibles). Les mondes virtuels sonores créés par la stéréophonie, la technologie VBAP ou encore par le binaural sur casque pourraient ainsi être comparés du point de vue de la *perception de la distance* et de la sensation de *présence* induites. Une autre perspective, axée sur l'étude du phénomène d'intégration spatiale audio-visuelle, est aussi envisageable. Dans l'étude de ce type de phénomènes, deux approches

sont possibles pour générer des incohérences spatiales entre les stimuli auditif et visuel [Godfroy et al. 2011] : la première consiste à créer physiquement ou virtuellement des disparités entre les stimuli auditif et visuel, la seconde consiste à proposer des stimuli auditifs et visuels éventuellement physiquement spatialement cohérents tout en s'assurant au préalable qu'ils ne sont pas perçus individuellement aux mêmes positions dans l'espace. Les résultats de l'expérience menée à la section 8.2 indiquent que dans notre cas la seconde approche n'est pas envisageable pour étudier l'intégration spatiale audio-visuelle en distance (voir la discussion de la section 8.3.1). L'intégration spatiale audio-visuelle en distance pourrait donc être étudiée ici en créant virtuellement des disparités physiques entre les stimuli auditif et visuel et en étudiant la position audio-visuelle perçue résultant de ces disparités.

Chapitre 9

Conclusion générale

Sommaire

9.1 Contributions de la thèse	235
9.1.1 Son spatialisé et réalité virtuelle	236
9.1.2 Comportement dynamique des LaMAPs comme sources sonores pour l'holophonie	236
9.1.3 Perception spatiale du monde virtuel synthétisé par le SMART-I ²	237
9.2 Quelques perspectives ...	238
9.2.1 Améliorations possibles du SMART-I ²	238
9.2.2 Impact d'un rendu sonore spatialisé sur des applications de réalité virtuelle	238
9.2.3 Non-linéarités générées par un système vibrant	239
9.2.4 Effets des inhomogénéités introduites par les nids d'abeille	239
9.2.5 Modélisation des LaMAPs en vue d'un contrôle vibratoire pour l'holophonie	240
9.2.6 Variations autour du protocole expérimental de la section 8.2	240
9.3 Publications liées à la thèse	241
9.3.1 Articles	241
9.3.2 Acte de conférences	241

La problématique étudiée tout au long de ce manuscrit était : comment transformer des *vibrations* en indices sonores spatiaux *perceptifs* d'un *environnement virtuel* ? Les contributions apportées par cette thèse à cette problématique sont récapitulées dans la section 9.1. En quelques mots, il est établi que le SMART-I² permet à un utilisateur d'utiliser ses capacités perceptives audio-visuelles et ses capacités motrices pour comprendre l'espace virtuel, mais qu'il est fondamentalement limité par ses imperfections dans sa quête de mondes virtuels procurant un sentiment de *présence* toujours plus grand. Ces limitations ont une conséquence majeure sur la perception de l'environnement virtuel : le monde virtuel n'est pas encore l'égal du monde réel. Ceci n'est, heureusement, pas une surprise et laisse donc ouvert de nombreuses perspectives, quelques unes d'entre elles étant présentées à la section 9.2.

9.1 Contributions de la thèse

Les contributions apportées par cette thèse à la problématique mentionnée précédemment sont analysées dans cette section selon les trois axes principaux définis dans l'introduction (chapitre 1).

9.1.1 Son spatialisé et réalité virtuelle

Pourquoi et comment intégrer un rendu sonore spatialisé dans un système de type “CAVE” ?

Un *rendu sonore spatialisé* est caractérisé par quatre attributs : la *direction*, la *distance*, la *réverbération* et l'*effet de parallaxe*. L'étude bibliographique menée à la section 2.3 montre que les différents attributs d'un *rendu sonore spatialisé* ont un impact considérable sur un grand nombre de facteurs importants dans les applications de réalité virtuelle : présence, intelligibilité et qualité audio-visuelle globale, par exemple. De plus, la cohérence audio-visuelle spatiale semble jouer un rôle important dans cet impact. Cette étude justifie donc de chercher à ajouter un *rendu sonore spatialisé* à un système destiné à la réalité virtuelle.

Le SMART-I² propose alors une intégration originale d'un système de *rendu sonore spatialisé* de type *holophonique* au sein d'un système de rendu *visuel* basé sur des écrans de grandes dimensions. Le schéma d'intégration proposé à la section 2.5 résout de plus deux problèmes pratiques majeurs : les conflits de positionnement entre les écrans et les haut-parleurs, et le filtrage du champ sonore induit par les écrans de projection.

Comment évaluer l'impact de l'ajout d'un système de rendu sonore spatialisé ?

L'impact des différents attributs du son spatialisé sur la sensation de *présence* dans des applications de réalité virtuelle peut être étudié grâce au SMART-I². Deux projets en cours et dont c'est l'objectif sont présentés : GAVIP (voir section 3.2) et MARC & SMART-I² (voir section 3.3). GAVIP propose une immersion *active* des participants dans le monde virtuel, tandis que l'autre projet propose une immersion plutôt *passive*. GAVIP n'a pas encore été évalué par les utilisateurs et ne fournit donc pour l'instant pas de résultats relativement aux objectifs mentionnés précédemment. Le projet MARC & SMART-I² a permis une évaluation assez sommaire qui permet de supposer que la cohérence audio-visuelle et l'interaction avec un personnage virtuel permettent d'augmenter la sensation de *présence*.

9.1.2 Comportement dynamique des LaMAPs comme sources sonores pour l'holophonie

Comment caractériser les non-linéarités générées par un système vibro-acoustique ?

L'estimation des non-linéarités générées par les systèmes vibrants (haut-parleurs, pots vibrants, etc.) a été abordée via la mise au point théorique et la validation numérique et expérimentale d'une méthode originale, dérivée des travaux empiriques d'Angelo Farina [Farina 2000, Farina et al. 2001]. Cette méthode permet d'estimer, au moyen d'une seule mesure, un modèle de système non-linéaire correspondant à un système vibrant (voir section 4.2). En complément, l'influence de différents paramètres sur la qualité des noyaux estimés est analysée et le cas d'une chaîne de systèmes non-linéaires est étudié. Une boîte à outils Matlab proposant une implémentation de cette méthode est disponible en ligne.

Comment modéliser le comportement dynamique d'un matériau sandwich ?

Une méthode permettant l'estimation des paramètres mécaniques nécessaires à la description de la dynamique des panneaux de matériau sandwich composant le SMART-I² est présentée et validée (voir section 5.3). Cette méthode ne nécessite qu'une seule mesure dynamique, ponctuelle, réalisée sur une plaque

taillée dans le matériau à décrire. Elle permet l'estimation des paramètres d'élasticité et d'amortissement du matériau testé, relatifs à une dynamique de plaque. Ces paramètres sont estimés par le biais d'une procédure mixte numérique/expérimentale consistant à mettre à jour les paramètres d'un modèle numérique en cherchant à minimiser l'écart entre les données obtenues expérimentalement et les données prédites par ce modèle. De plus, les hypothèses de *fluide léger*, de *faible amortissement* et de *pertes par rayonnement négligeables*, faites pour les panneaux étudiés expérimentalement, sont validés.

Quelle est, en pratique, le comportement dynamique d'un LaMAP ?

Différents phénomènes influençant le comportement dynamique du SMART-I² (les conditions aux limites, l'influence des excitateurs les uns sur les autres, les interactions ayant lieu entre les deux LaMAPs composant le SMART-I²) sont décrits en confrontant les résultats expérimentaux et les prédictions faites par un modèle simple (modèle de plaque mince, homogène, isotrope, infinie et excitation ponctuelle) au chapitre 6. Cette étude révèle que le LaMAP étudié ne possède pas des conditions aux limites homogènes. De plus, les effets des excitateurs voisins sur le champ de vitesse généré par un excitateur donné induisent des interférences temporelles entre le champ direct et le champ réfléchi par les excitateurs voisins qui mettent en défaut le modèle théorique simple utilisé. En ce qui concerne l'hypothèse d'excitation ponctuelle, les résultats expérimentaux suggèrent qu'elle est raisonnable en basses fréquences.

9.1.3 Perception spatiale du monde virtuel synthétisé par le SMART-I²

Le monde virtuel audio-visuel créé par le SMART-I² est-il perceptivement spatialement cohérent ?

Deux expériences cherchent à estimer la précision de localisation en azimuth, la capacité de restitution de l'effet de parallaxe auditive, ainsi que la cohérence perceptive du rendu audio-visuel proposé par le SMART-I² (section 7.4). Le dispositif génère un rendu audio spatialisé permettant une localisation en azimuth aussi précise, mais légèrement plus variable, que dans la réalité ou qu'avec un système de WFS conventionnel. Le système procure à ses utilisateurs une sensation de distance réaliste, mais compressée, en utilisant l'effet de parallaxe auditive. L'environnement virtuel proposé par le SMART-I² présente suffisamment de cohérence audio-visuelle pour que les sujets soient à même d'intégrer les flux auditif et visuel qui leur sont présentés. Le SMART-I² permet donc à un utilisateur, plongé dans l'environnement de réalité virtuelle qu'il restitue, d'utiliser ses capacités perceptives audio-visuelles et ses capacités motrices pour comprendre l'espace virtuel et se déplacer en son sein.

Quel est l'impact de la transparence limitée du SMART-I² sur l'espace virtuel perçu ?

Une expérience explorant l'influence de la *transparence* limitée du SMART-I² sur la perception de la distance et sur la sensation de *présence* a été réalisée (section 8.2). Du point de vue *perceptif* cette expérience montre que, de façon analogue à l'expérience présentée à la section 7.4.4 et à la littérature, les distances perçues dans l'environnement virtuel sont systématiquement sous-estimées pour des distances de rendu inférieures à 2.3 m et sur-estimées pour des distances de rendu supérieures à 2.3 m. Étonnamment, les distances sont perçues de façon similaire quelle que soit la combinaison de modalités impliquée (audio, visuel ou audio-visuel). Du point de vue d'un éventuel lien entre *perception* et *cognition*, la sous-estimation

des distances est liée à la sensation de *présence* pour la majorité des sujets : plus le sujet se sent *présent* dans le monde virtuel, moins il sous-estime les distances dans le monde virtuel. D'autre part, le système de restitution audio-visuelle n'est pas complètement *transparent* d'un point de vue perceptif : il est supposé que de par sa présence physique, le SMART-I² *ancree* le monde virtuel dans le monde réel, à la fois pour la modalité auditive (réverbération de la salle d'écoute) et la modalité visuelle (accommodation à la distance de l'écran).

9.2 Quelques perspectives ...

Quelques perspectives ouvertes par cette thèse sont proposées dans cette section.

9.2.1 Améliorations possibles du SMART-I²

Parmi les améliorations possibles du SMART-I², il est possible de mentionner que le “M” du SMART-I², faisant référence à l'aspect multi-utilisateur, n'est en réalité que partiellement opérationnel. La stéréoscopie passive avec suivi (section 2.5.5) ne permet qu'à un seul utilisateur de bénéficier de l'environnement visuel virtuel. Le rendu sonore spatialisé restitué par le SMART-I², car il est basé sur une reconstruction physique du champ sonore, permet par contre à plusieurs utilisateurs de bénéficier simultanément de l'environnement audio virtuel. Une façon de proposer un rendu visuel à deux utilisateurs consisterait à utiliser les filtres polarisants en place pour envoyer aux deux yeux de chaque utilisateur l'image monoscopique de la scène correspondant à leur position. Une telle solution, que l'on pourrait qualifier de *monoscopie passive avec suivi*, mérite donc d'être étudiée.

Une autre amélioration possible du dispositif concerne la restitution de l'élévation des objets sonores. En effet, la “Wave Field Synthesis” étant basée sur un banc de haut-parleurs répartis sur une ligne horizontale (voir section 2.5.3), l'élévation n'est pas restituée. Toutefois, la première question à laquelle il faut répondre est : “Est-il, dans des applications audio-visuelles de réalité virtuelle, nécessaire de restituer l'élévation des objets sonores virtuels ?”. Les analyses bibliographiques des sections 7.3.1 et 2.3 suggèrent que c'est possiblement le cas, mais que cela mérite d'être étudié de façon rigoureuse. Plusieurs solutions technologiques peuvent alors être envisagées. Une première solution serait de coller au dos des LaMAPs plusieurs lignes d'excitateurs électrodynamiques et de sélectionner la (ou les) lignes à piloter en fonction de l'élévation de la source. Une seconde solution consisterait à simuler l'élévation des objets sonores virtuels par un filtrage fréquentiel similaire à celui qui serait effectué par le pavillon des oreilles à partir de l'analyse d'une base de donnée d'HRTFs [Lopez et al. 2010].

9.2.2 Impact d'un rendu sonore spatialisé sur des applications de réalité virtuelle

Comme évoqué dans les sections 2.3.2 et 2.3.3, l'ajout d'un *rendu sonore spatialisé* peut avoir un impact fort sur la sensation de *présence* ressentie par les utilisateurs du SMART-I². Cependant, comme le montre le tableau de synthèse 2.1 de la section 2.3.6, la nature de ces impacts reste encore mal connue, voire même inconnue dans certains cas. La plateforme mise en place par le projet GAVIP est fonctionnelle et deux scénarios ont été développés sur cette base. De plus, les premières réactions des utilisateurs suggèrent que l'intégration de la plateforme *MARC* au sein du SMART-I² est efficace. Enfin, un projet actuellement en

cours cherche à intégrer dans le SMART-I² un contenu visuel issu du cinéma 3D en y associant un rendu sonore spatialisé restitué par “*Wave Field Synthesis*” [Evrard et al. 2011]. Les contenus développés par ces trois projets peuvent donc à présent servir de socles à des expériences perceptives visant à déterminer l’impact des différents attributs d’un *rendu sonore spatialisé* dans des applications de réalité virtuelle.

9.2.3 Non-linéarités générées par un système vibrant

La mise au point et la validation de la méthode d’estimation des non-linéarités présentée à la section 4.2 répondent à l’interrogation “*Comment estimer les non-linéarités générées par un système vibrant ?*”. Cependant, les systèmes faiblement non-linéaires étudiés sont en pratique supposés être modélisables par des modèles de Hammerstein en cascade. Les résultats expérimentaux obtenus montrent que cette hypothèse conduit à des prédictions d’assez bonne qualité. Il serait toutefois satisfaisant d’être capable de valider cette hypothèse de façon rigoureuse. De plus, cette méthode permet d’estimer des “*diagonales de noyaux de Volterra*” relativement à un système non-linéaire donné. Récemment, Torras-Rosell & Jacobsen [2011] ont fait apparaître que les parties non-diagonales des noyaux de Volterra *polluent* également de façon anticausale le résultat de la déconvolution. Il est alors tentant de se demander s’il existe des signaux, multi-fréquentiels à un instant t donné, permettant l’estimation simultanée de plusieurs *tranches* d’un noyau de Volterra, et non plus seulement de la *tranche diagonale*. Enfin, les résultats présentés à la section 4.2 et ceux de Novák et al. [2010a] font apparaître qu’il existe au moins deux signaux différents possédant la propriété de phase fondamentale nécessaire à l’estimation des modèles de Hammerstein en cascade. Il serait alors intéressant de chercher mathématiquement la classe de signaux possédant cette propriété de phase fondamentale. Cela permettrait de choisir dans cette classe les signaux optimisant la qualité des estimations en fonction de certains critères (durée du signal d’entrée, rapport *signal sur bruit*, etc.).

9.2.4 Effets des inhomogénéités introduites par les nids d’abeille

En guise de perspective à la méthode d’estimation des paramètres mécaniques de panneaux de matériau sandwich décrite à la section 5.3, il est possible de chercher à étudier plus finement l’effet des inhomogénéités introduites par les nids d’abeille sur le comportement dynamique des panneaux étudiés ici. Tant que les longueurs d’ondes étudiées sont très grandes devant les dimensions des inhomogénéités, les vibrations «*ne voient pas les nids d’abeille*». Toutefois les nids d’abeille ont quand même tendance à induire un léger comportement non-déterministe dans la dynamique des panneaux étudiés à mesure que la fréquence augmente [Weaver 1989, Bohigas et al. 1991]. Ceci se traduit par le fait que des modes de fréquences propres proches auront une légère tendance à se *repousser* mutuellement. Les homogénéités ont aussi un effet sur la distribution statistique des amortissements modaux : ces effets n’ont cependant été étudiés pour l’instant que d’un point de vue numérique [Burkhardt & Weaver 1996].

Il est ici assez délicat de faire des statistiques concernant ce point étant donné le peu d’informations modales disponibles. Une perspective intéressante pourrait alors être d’essayer d’obtenir, grâce à l’analyse modale haute-résolution [Ege et al. 2009], plus de données expérimentales (fréquences modales et amortissement modaux) sur des panneaux contenant des inhomogénéités. Ceci permettrait d’évaluer expérimentalement l’influence des inhomogénéités sur les répartitions des écarts inter-modaux et des amortissements modaux et de comparer les résultats expérimentaux, numériques et théoriques.

9.2.5 Modélisation des LaMAPs en vue d'un contrôle vibratoire pour l'holophonie

Le chapitre 6 est une première tentative de modélisation dynamique réaliste des LaMAPs. Cette étude expérimentale a permis de juger de la capacité d'un modèle de plaque infinie à prédire le comportement dynamique des LaMAPs. Même si le recouvrement modal est important dans le domaine des fréquences audibles, les conditions aux limites interviennent de manière forte sur la dynamique des panneaux. Il semble donc approprié de chercher à modéliser les LaMAPs plutôt à l'aide d'un modèle statistique. Cependant, les approches statistiques habituelles (comme la SEA¹ par exemple) sont énergétiques et prédisent les résultats *en moyenne*. Or pour l'application visée ici, *i.e.* l'holophonie, les détails du comportement dynamique nous intéressent aussi "*dans une certaine mesure*" (qu'il conviendra de déterminer). Une perspective de ce point de vue là serait d'arriver à trouver un descripteur de la dynamique des LaMAPs qui soit adapté à l'objectif visé.

L'intérêt de disposer d'un modèle permettant de prédire la dynamique des LaMAPs est considérable. En effet, la connaissance de la dynamique de ces panneaux permet d'envisager un contrôle direct du comportement vibratoire des LaMAPs. Le problème de l'holophonie pourrait ainsi être abordé, non plus d'un point de vue acoustique, mais d'un point de vue vibratoire. Un rendu sonore spatialisé holophonique pourrait alors être généré en synthétisant directement sur le panneau le champ de vitesse correspondant à une source sonore virtuelle placée à une position donnée dans l'espace. Cette procédure aurait de plus l'avantage d'être moins sensible à l'influence acoustique de la zone de restitution qu'une procédure uniquement basée sur un contrôle acoustique.

9.2.6 Variations autour du protocole expérimental de la section 8.2

Le protocole expérimental développé pour l'expérience décrite à la section 8.2 autorise de nombreuses variations permettant d'aborder de nouvelles perspectives.

La première perspective proposée ici concerne la comparaison de différents types de rendu sonores spatialisés. L'expérience décrite à la section 8.2 a été menée en utilisant la "*Wave Field Synthesis*" comme système de rendu sonore spatialisé. Il est cependant possible, et même assez aisé, de réaliser la même expérience en incluant le système de rendu sonore spatialisé comme variable indépendante dans le protocole expérimental. Les mondes virtuels sonores créés par la stéréophonie, la technologie VBAP ou encore par le binaural sur casque² pourraient ainsi être comparés du point de vue de la *perception de la distance* et de la sensation de *présence* induites.

Une autre perspective, axée sur l'étude du phénomène d'intégration spatiale audio-visuelle, est aussi envisageable. Deux approches sont possibles pour générer des incohérences spatiales entre les stimuli auditif et visuel [Godfroy et al. 2011] : la première consiste à créer physiquement ou virtuellement des disparités entre les stimuli auditif et visuel, la seconde consiste à proposer des stimuli auditifs et visuels éventuellement physiquement spatialement cohérents tout en s'assurant au préalable qu'ils ne sont pas perçus individuellement aux mêmes positions dans l'espace. Les résultats de l'expérience menée à la section 8.2 indiquent que dans notre cas la seconde approche n'est pas envisageable pour étudier l'intégration spatiale

1. *Statistical Energy Analysis*, voir [Bruneau & Scelo 2006] par exemple.

2. Pour un aperçu des différentes technologies de rendu sonore spatialisé disponibles, on se référera à la section 2.4.2 et à la table 2.2.

audio-visuelle en distance (voir la discussion dans la section 8.3.1). L'intégration spatiale audio-visuelle en distance pourrait donc être étudiée ici en créant virtuellement des disparités physiques entre les stimuli auditif et visuel et en étudiant la position audio-visuelle perçue résultant de ces disparités.

9.3 Publications liées à la thèse

Cette section regroupe les différentes publications qui sont liées au manuscrit présenté ici.

9.3.1 Articles

- **[Rébillat et al. 2011c]** : M. RÉBILLAT, R. HENNEQUIN, E. CORTEEL & B. F.G. KATZ
“Identification of cascade of Hammerstein models for the description of nonlinearities in vibrating devices”, *Journal of Sound and Vibration*, 2011, 330, 1018-1038.
- **[Rébillat & Boutillon 2011a]** : M. RÉBILLAT & X. BOUTILLON
“Measurement of relevant elastic and damping material properties in sandwich thick-plates”, accepté le 13/07/11 par le *Journal of Sound and Vibration*.
- **[Rébillat et al. 2011a]** : M. RÉBILLAT, X. BOUTILLON, E. CORTEEL & B. F.G. KATZ
“Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments”, soumis le 20/08/11 à *ACM Transactions on Applied Perception*.
- **[Ege et al. 2011]** : K. EGE, X. BOUTILLON & M. RÉBILLAT
“Vibroacoustics of the piano soundboard (part 1) : non-linearity and modal properties in the low- and mid-frequency ranges”, en préparation pour le *Journal of Sound and Vibration*.

9.3.2 Acte de conférences

- **[Rébillat et al. 2008]** : M. RÉBILLAT, E. CORTEEL & B. F.G. KATZ
“SMART-I² : Spatial Multi-user Audio-visual Real-Time Interactive Interface”, *125th Convention of the Audio Engineering Society*, 2008.
- **[Rébillat et al. 2009]** : M. RÉBILLAT, B. F.G. KATZ & E. CORTEEL
“SMART-I² : Spatial Multi-user Audio-visual Real-Time Interactive Interface, a Broadcast Application Context”, *Proceedings of the IEEE 3D-TV Conference*, 2009.
- **[Rébillat & Boutillon 2010a]** : M. RÉBILLAT & X. BOUTILLON
“Identification of elastic and damping properties of sandwich structures based on high resolution modal analysis of point measurements”, *ISMA*, Leuven, Belgium, 2010.
- **[Rébillat et al. 2010a]** : M. RÉBILLAT, E. CORTEEL & B. F.G. KATZ
“The SMART-I² : A new approach for the design of immersive audio-visual environments”, *Euro-VR Eve 2010*, Orsay, France, 2010.
- **[Rébillat et al. 2010b]** : M. RÉBILLAT, R. HENNEQUIN, E. CORTEEL & B. F.G. KATZ
“Prediction of harmonic distortion generated by electro-dynamic loudspeakers using cascade of Hammerstein models”, *128th Convention of the Audio Engineering Society*, 2010.
- **[Rébillat & Boutillon 2010b]** : M. RÉBILLAT & X. BOUTILLON
“Identification of honeycomb sandwich properties by high-resolution modal analysis”, *ICEM*, Poitiers,

France, 2010.

- **[Courgeon et al. 2010]** : M. COURGEON, M. RÉBILLAT, B. F.G. KATZ, C. CLAVEL & J.-C. MARTIN
 “Life-Sized Audiovisual Spatial Social Scenes with Multiple Characters : MARC & SMART-I²”
5ièmes journées de l’Association Française de Réalité Virtuelle (AFRV), 2010.
- **[Rébillat & Boutillon 2011b]** : M. RÉBILLAT & X. BOUTILLON
 “A method to measure elastic and dissipative material properties of sandwich structures and its numerical validation”, *CSMA*, Giens, France, 2011.
- **[Rébillat et al. 2011b]** : M. RÉBILLAT, X. BOUTILLON, E. CORTEEL & B. F.G. KATZ
 “Audio, visual and audio-visual egocentric distance perception in virtual environments”, *Proceedings of the Forum Acusticum*, Aalborg, Danemark, 2011.
- **[Caramiaux et al. 2011]** : B. CARAMIAUX, S. FDILI ALAOU, T. BOUCHARA, G. PARSEIHIAN, & M. RÉBILLAT
 “Gestural auditory and visual interactive platform” *Proceedings of the 14th International Conference on Digital Audio Effects (DAFX-11)*, Paris, France, 2011.

Annexe A

Annexe “Non-linéarités” : Calcul de la matrice C

Les polynômes de Tchebychev $\{T_k[\cos(\phi)]\}_{k \in \mathbb{N}}$ sont définis par l'équation (A.1).

$$\forall k \in \mathbb{N}, \cos(k\phi) = T_k[\cos(\phi)] \quad (\text{A.1})$$

Il peut facilement être montré qu'ils satisfont à la relation de récurrence donnée par l'équation (A.2).

$$k = 0 \quad T_0(x) = 1 \quad (\text{A.2a})$$

$$k = 1 \quad T_1(x) = x \quad (\text{A.2b})$$

$$k > 1 \quad T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x) \quad (\text{A.2c})$$

En utilisant la forme suggérée par l'équation A.3 pour écrire les polynômes et en reportant ceux-ci dans l'équation (A.2), l'équation (A.4) est obtenue. Cette équation permet de déduire les coefficients de la matrice A .

$$T_k(x) = \sum_{i=0}^k A(i, k) x^i \quad (\text{A.3})$$

$$i = 0 \quad A(0, k+1) = -A(0, k-1) \quad (\text{A.4a})$$

$$0 < i < k \quad A(i, k+1) = 2A(i-1, k) - A(i, k-1) \quad (\text{A.4b})$$

$$i \geq k \quad A(i, k+1) = 2A(i-1, k) \quad (\text{A.4c})$$

La linéarisation des polynômes peut être réécrite sous forme matricielle :

$$\begin{bmatrix} 1 \\ \cos(x) \\ \dots \\ \cos(Nx) \end{bmatrix} = \mathbf{A} \begin{bmatrix} 1 \\ \cos(x) \\ \dots \\ \cos^N(x) \end{bmatrix} \quad (\text{A.5})$$

En inversant l'équation (A.5), l'équation (4.7) est obtenue directement. L'équation (A.6) donne ainsi explicitement la matrice C .

$$\mathbf{C} = \mathbf{A}^{-1} \quad (\text{A.6})$$

La matrice \mathbf{A}_c^T , nécessaire pour accéder aux $\{h_n(t)\}_{n \in [1, N]}$, est la matrice \mathbf{A} sans la première colonne ni la première ligne, comme le montre l'équation (4.18). Pour éviter l'implémentation de la récurrence, la transposée de la matrice \mathbf{A}_c est donnée à l'ordre 8 par l'équation (A.7), ce qui est suffisant en pratique.

$$\mathbf{A}_c^T = \begin{bmatrix} 1 & 0 & -3 & 0 & 5 & 0 & -7 & 0 \\ 0 & 2 & 0 & -8 & 0 & 18 & 0 & -32 \\ 0 & 0 & 4 & 0 & -20 & 0 & 56 & 0 \\ 0 & 0 & 0 & 8 & 0 & -48 & 0 & 160 \\ 0 & 0 & 0 & 0 & 16 & 0 & -112 & 0 \\ 0 & 0 & 0 & 0 & 0 & 32 & 0 & -256 \\ 0 & 0 & 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 128 \end{bmatrix} \quad (\text{A.7})$$

Annexe B

Détails concernant la procédure d'estimation des propriétés d'élasticité et de dissipation des constituants de plaques “sandwich”

B.1 Génération de la base polynomiale orthonormale

La base orthonormale polynomiale $\{p_i(\kappa)\}_{i \in [0, N]}$ [Bhat 1985, Kim 1986] utilisée dans la procédure de Rayleigh étendue présentée à la section 5.3.3 est générée par un processus itératif de Gram-Schmidt de la façon suivante :

$$\begin{cases} i = 0 & \tilde{p}_0(\kappa) = \frac{1}{\sqrt{2}} \\ i = 1 & \tilde{p}_1(\kappa) = (\kappa - \langle \kappa p_0(\kappa), p_0(\kappa) \rangle) p_0(\kappa) \\ i > 1 & \tilde{p}_i(\kappa) = (\kappa - \langle \kappa p_{i-1}(\kappa), p_{i-1}(\kappa) \rangle) p_{i-1}(\kappa) - \langle \kappa p_{i-2}(\kappa), p_{i-2}(\kappa) \rangle p_{i-2}(\kappa) \end{cases} \quad (\text{B.1})$$

Le produit scalaire entre deux polynômes $a(\kappa)$ et $b(\kappa)$ est : $\langle a(\kappa), b(\kappa) \rangle = \int_{-1}^1 a(\kappa) b(\kappa) d\kappa$. Les versions normalisées et non-normalisées du $i^{\text{ième}}$ élément de la base polynômiale sont notées respectivement $p_i(\kappa)$ et $\tilde{p}_i(\kappa)$ et sont liées par la relation : $\tilde{p}_i(\kappa) = \sqrt{\langle \tilde{p}_i(\kappa), \tilde{p}_i(\kappa) \rangle} p_i(\kappa)$. L'équation suivante étant satisfaite, la base est orthonormale : $\forall (i, j) \in [0, N]^2 \quad \langle p_i(\kappa), p_j(\kappa) \rangle = \delta_{ij}$, où δ_{ij} est le symbole de Kronecker.

B.2 Expression analytiques de T et U

Les matrices K et M du problème aux valeurs propres $[K - 4\pi^2 f^2 M]q = 0$ (Sec. 5.3.3) sont issues des expressions analytiques des dérivées de T et U relativement aux déplacements généralisés L_{ij} , M_{ij} , N_{ij} . Ces déplacements généralisés sont liés aux déplacements “naturels” Φ_x , Φ_y , et w_0 par ;

$$\Phi_x(x, y) = \sum_{i,j} L_{ij} p_i(x) p_j(y) \quad \Phi_y(x, y) = \sum_{i,j} M_{ij} p_i(x) p_j(y) \quad w_0(x, y) = \sum_{i,j} N_{ij} p_i(x) p_j(y) \quad (\text{B.2})$$

Les dérivées de l'énergie cinétique $T = \frac{\rho^H \omega^2}{2} \iint_{(s)} \left[\frac{h^3}{12} (\Phi_x^2 + \Phi_y^2) + h w_0^2 \right] dx dy$ sont :

$$\begin{cases} \frac{\partial T}{\partial L_{ij}} = \frac{\rho^H l_x l_y h \omega^2}{4} L_{ij} \\ \frac{\partial T}{\partial M_{ij}} = \frac{\rho^H l_x l_y h \omega^2}{4} M_{ij} \\ \frac{\partial T}{\partial N_{ij}} = \frac{\rho^H l_x l_y h \omega^2}{4} N_{ij} \end{cases} \quad (\text{B.3})$$

La simplicité de ces formules est due au fait qu'aucune dérivée spatiale n'apparaît dans l'expression de l'énergie cinétique et au fait que tout les produits $p_i p_j$ ($i \neq j$) s'annulent une fois intégrés par orthogonalité des polynômes.

Les dérivées de l'énergie potentielle U (donnée par l'Eq. 5.13) sont :

$$\begin{cases} \frac{\partial U}{\partial L_{ij}} = \sum_{k=0}^N \left[-\frac{l_x}{2} D_4 N_{ik} I(j, k) + \frac{l_x}{l_y} D_4 L_{ik} J(j, k) - \frac{l_y}{2} D_5 M_{kj} I(i, k) + \frac{l_y}{l_x} D_5 L_{kj} J(i, k) \right] \\ \frac{\partial U}{\partial M_{ij}} = \frac{l_x l_y}{4} D_5 M_{ij} + \sum_{k=0}^N \left[\frac{l_y}{l_x} D_1 M_{kj} J(i, k) - \frac{l_y}{2} D_5 L_{kj} I(k, i) + \frac{l_x}{l_y} D_6 M_{ik} J(j, k) \right] + \dots \\ \dots \sum_{k,l} \left[\frac{D_2}{2} I(i, k) I(l, j) + D_6 I(k, i) I(j, l) \right] N_{kl} \\ \frac{\partial U}{\partial N_{ij}} = \frac{l_x l_y}{4} D_4 N_{ij} + \sum_{k=0}^N \left[\frac{l_x}{l_y} D_3 N_{ik} J(k, j) - \frac{l_x}{2} D_4 L_{ik} I(k, j) + \frac{l_y}{l_x} D_6 N_{kj} J(k, i) \right] + \dots \\ \dots \sum_{k,l} \left[\frac{D_2}{2} I(k, i) I(j, l) + D_6 I(i, k) I(l, j) \right] M_{kl} \end{cases} \quad (\text{B.4})$$

où $\sum_{k,l}$ remplace $\sum_{(k,l) \in [0,N]^2}$ et où les intégrales suivantes ont été introduites :

$$I(i, j) = \left\langle \frac{dp_i(\kappa)}{d\kappa}, p_j(\kappa) \right\rangle = \int_{-1}^1 \frac{dp_i(\kappa)}{d\kappa} p_j(\kappa) d\kappa \quad (\text{B.5})$$

$$J(i, j) = \left\langle \frac{dp_i(\kappa)}{d\kappa}, \frac{dp_j(\kappa)}{d\kappa} \right\rangle = \int_{-1}^1 \frac{dp_i(\kappa)}{d\kappa} \frac{dp_j(\kappa)}{d\kappa} d\kappa \quad (\text{B.6})$$

B.3 Expressions analytiques de t_n et u_k^n

Les expressions analytiques des coefficients t_n et u_k^n , représentant respectivement la n -ième masse modale (pour un déplacement normalisé donné) et la k -ième contribution à la n -ième rigidité modale, sont nécessaires pour obtenir numériquement les amortissements modaux et les sensibilités des fréquences

et amortissements modaux aux rigidités et aux facteurs d'amortissement, comme expliqué dans les sections 5.3.3 et 5.3.5. Les coordonnées de la $n^{\text{ième}}$ déformée modale sont notées $\{L_{lm}^n, M_{lm}^n, N_{lm}^n\}$. Les calculs sont effectués à la section 5.3.3. L'expression de t_n est :

$$t_n = \frac{\rho^H l_x l_y h}{8} \sum_{l,m} \left[(L_{lm}^n)^2 + \frac{h^2}{12} ((M_{lm}^n)^2 + (N_{lm}^n)^2) \right] \quad (\text{B.7})$$

Les expressions des $\{u_k^n\}_{k \in [1,6]}$ sont :

$$\left\{ \begin{array}{l} u_1^n = \frac{l_y}{2l_x} \sum_{l,m} M_{lm}^n \left[\sum_{p=0}^N M_{pm}^n J(l, p) \right] \\ u_2^n = \frac{1}{2} \sum_{l,m} M_{lm}^n \left[\sum_{p,q} N_{pq}^n I(l, p) I(l, q) \right] \\ u_3^n = \frac{l_x}{2l_y} \sum_{l,m} N_{ml}^n \left[\sum_{p=0}^N N_{lp}^n J(m, p) \right] \\ u_4^n = \frac{l_x l_y}{8} \sum_{l,m} (N_{ml}^n)^2 - \frac{l_x}{2} \sum_{l,m} N_{ml}^n \left[\sum_{p=0}^N L_{lp}^n I(p, m) \right] + \frac{l_x}{2l_y} \sum_{l,m} L_{ml}^n \left[\sum_{p=0}^N L_{lp}^n J(m, p) \right] \\ u_5^n = \frac{l_x l_y}{8} \sum_{l,m} (M_{ml}^n)^2 - \frac{l_y}{2} \sum_{l,m} M_{ml}^n \left[\sum_{p=0}^N L_{pm}^n I(p, l) \right] + \frac{l_y}{2l_x} \sum_{l,m} L_{ml}^n \left[\sum_{p=0}^N L_{pm}^n J(l, p) \right] \\ u_6^n = \frac{l_x}{2l_y} \sum_{l,m} M_{ml}^n \left[\sum_{p=0}^N M_{lp}^n J(m, p) \right] + \sum_{l,m} M_{ml}^n \left[\sum_{p,q} N_{pq}^n I(p, l) I(m, q) \right] + \dots \\ \dots + \frac{l_y}{2l_x} \sum_{l,m} N_{ml}^n \left[\sum_{p=0}^N N_{pm}^n J(l, p) \right] \end{array} \right. \quad (\text{B.8})$$

où $I(l, p)$ et $J(l, p)$ sont définies à la fin de la section B.2.

B.4 Nomenclature de la section 5.3

l_x	(m)	Longueur du côté selon x de la plaque
l_y	(m)	Longueur du côté selon y de la plaque
h^s	(m)	Épaisseur de la peau
h^c	(m)	Épaisseur du coeur
$h = h^c + 2h^s$	(m)	Épaisseur du panneau sandwich
ρ^s	(kg × m ⁻³)	Masse volumique de la peau
ρ^c	(kg × m ⁻³)	Masse volumique du coeur
$\rho^H = 1/h \times (h^c \rho^c + 2h^s \rho^s)$	(kg × m ⁻³)	Masse volumique de la plaque homogène équivalente
s_{cell}	(m)	Longueur caractéristique d'une cellule de nids d'abeille

TAB. B.1 – Notations utilisées pour les paramètres massiques et géométriques des panneaux.

	Coeur	Peau	Plaque homogène équivalente
Module de Young dans la direction x	$\underline{E}_x^c = E_x^c(1 + j\eta_x^c)$	$\underline{E}_x^s = E_x^s(1 + j\eta_x^s)$	$\underline{E}_x^H = E_x^H(1 + j\eta_x^H)$
Module de Young dans la direction y	$\underline{E}_y^c = E_y^c(1 + j\eta_y^c)$	$\underline{E}_y^s = E_y^s(1 + j\eta_y^s)$	$\underline{E}_y^H = E_y^H(1 + j\eta_y^H)$
Module de Young dans la direction z	$\underline{E}_z^c = E_z^c(1 + j\eta_z^c)$	$\underline{E}_z^s = E_z^s(1 + j\eta_z^s)$	$\underline{E}_z^H = E_z^H(1 + j\eta_z^H)$
Module de cisaillement dans le plan xy	$\underline{G}_{xy}^c = G_{xy}^c(1 + j\eta_{xy}^c)$	$\underline{G}_{xy}^s = G_{xy}^s(1 + j\eta_{xy}^s)$	$\underline{G}_{xy}^H = G_{xy}^H(1 + j\eta_{xy}^H)$
Coefficient de Poisson dans le plan xy	$\underline{\nu}_{xy}^c = \nu_{xy}^c(1 + j\eta_{\nu_{xy}}^c)$	$\underline{\nu}_{xy}^s = \nu_{xy}^s(1 + j\eta_{\nu_{xy}}^s)$	$\underline{\nu}_{xy}^H = \nu_{xy}^H(1 + j\eta_{\nu_{xy}}^H)$
Module de cisaillement dans le plan xz	$\underline{G}_{xz}^c = G_{xz}^c(1 + j\eta_{xz}^c)$	$\underline{G}_{xz}^s = G_{xz}^s(1 + j\eta_{xz}^s)$	$\underline{G}_{xz}^H = G_{xz}^H(1 + j\eta_{xz}^H)$
Coefficient de Poisson dans le plan xz	$\underline{\nu}_{xz}^c = \nu_{xz}^c(1 + j\eta_{\nu_{xz}}^c)$	$\underline{\nu}_{xz}^s = \nu_{xz}^s(1 + j\eta_{\nu_{xz}}^s)$	$\underline{\nu}_{xz}^H = \nu_{xz}^H(1 + j\eta_{\nu_{xz}}^H)$
Module de cisaillement dans le plan yz	$\underline{G}_{yz}^c = G_{yz}^c(1 + j\eta_{yz}^c)$	$\underline{G}_{yz}^s = G_{yz}^s(1 + j\eta_{yz}^s)$	$\underline{G}_{yz}^H = G_{yz}^H(1 + j\eta_{yz}^H)$
Coefficient de Poisson dans le plan yz	$\underline{\nu}_{yz}^c = \nu_{yz}^c(1 + j\eta_{\nu_{yz}}^c)$	$\underline{\nu}_{yz}^s = \nu_{yz}^s(1 + j\eta_{\nu_{yz}}^s)$	$\underline{\nu}_{yz}^H = \nu_{yz}^H(1 + j\eta_{\nu_{yz}}^H)$

TAB. B.2 – Notations utilisées pour les paramètres mécaniques complexes (paramètres élastiques et paramètres d'amortissement) des panneaux ($j^2 = -1$)

$u(x, y, z)$	(m)	Déplacement dans la direction x
$v(x, y, z)$	(m)	Déplacement dans la direction y
$w(x, y, z)$	(m)	Déplacement dans la direction z
$\Phi_x(x, y)$	(rad)	Rotation autour de l'axe x (hypothèse de Reissner-Mindlin)
$\Phi_y(x, y)$	(rad)	Rotation autour de l'axe y (hypothèse de Reissner-Mindlin)
$w_0(x, y)$	(m)	Déplacement dans la direction z (hypothèse de Reissner-Mindlin)
U	(J)	Énergie potentielle
ΔU	(J)	Énergie perdue par cycle
T	(J)	Énergie cinétique
$D_1 = \frac{E_x^H h^3}{12(1 - \nu_{xy}\nu_{yx})}$	(N×m)	Rigidité de la plaque dans la direction x
$D_2 = \frac{\nu_{xy} E_y^H h^3}{6(1 - \nu_{xy}\nu_{yx})}$	(N×m)	Rigidité de la plaque dans le plan xy
$D_3 = \frac{E_y^H h^3}{12(1 - \nu_{xy}\nu_{yx})}$	(N×m)	Rigidité de la plaque dans la direction y
$D_4 = 2\kappa_{yz}^2 h G_{yz}^H$	(N×m)	Rigidité de la plaque dans le plan yz
$D_5 = 2\kappa_{xz}^2 h G_{xz}^H$	(N×m)	Rigidité de la plaque dans le plan xz
$D_6 = \frac{G_{xy}^H h^3}{6}$	(N×m)	Rigidité de la plaque dans le plan xy
κ_{xz}	-	Facteur de correction pour le cisaillement dans le plan xz
κ_{yz}	-	Facteur de correction pour le cisaillement dans le plan yz
$\eta_\nu^{c,s,H} = \eta_{\nu_{xy}}^{c,s,H} + \eta_y^{c,s,H} = \eta_{\nu_{yx}}^{c,s,H} + \eta_x^{c,s,H}$	-	Facteur de perte global dû aux effets du coefficient de Poisson dans le plan xy
$\eta_1 = \eta_x^H$	-	Facteur de perte dans la direction x
$\eta_2 = \eta_\nu^H$	-	Facteur de perte global dû aux effets du coefficient de Poisson dans le plan xy
$\eta_3 = \eta_y^H$	-	Facteur de perte dans la direction y
$\eta_4 = \eta_{yz}^H$	-	Facteur de perte dans le plan yz
$\eta_5 = \eta_{xz}^H$	-	Facteur de perte dans le plan xz
$\eta_6 = \eta_{xy}^H$	-	Facteur de perte dans le plan xy
f	(Hz)	Fréquence
$\omega = 2\pi f$	(rad×s ⁻¹)	Pulsation
λ	(m)	Longueur d'onde des ondes de flexion
λ_x^{\min}	(m)	Longueur d'onde minimale des ondes de flexion dans la direction x dans le domaine fréquentiel étudié
λ_y^{\min}	(m)	Longueur d'onde minimale des ondes de flexion dans la direction y dans le domaine fréquentiel étudié

TAB. B.3 – Notations utilisée pour la modélisation dynamique des plaques

K	$(N \times m^{-1})$	Matrice de rigidité
M	(kg)	Matrice de masse
C	$(N \times m^{-1} \times s^{-1})$	Matrice d'amortissement
f_n^{XP}	(Hz)	Fréquence modale expérimentale du $n^{ième}$ mode
α_n^{XP}	(s^{-1})	Amortissement modal expérimentale du $n^{ième}$ mode
f_n^{Num}	(Hz)	Fréquence modale numérique du $n^{ième}$ mode
α_n^{Num}	(s^{-1})	Amortissement modal numérique du $n^{ième}$ mode
ξ_n^{Num}	(m)	Déformée modale numérique du $n^{ième}$ mode
μ	-	Recouvrement modal
\mathcal{P}_C	-	Système conservatif associé à la plaque
U_n^C	(J)	Énergie potentielle du $n^{ième}$ mode de \mathcal{P}_C
T_n^C	(J)	Énergie cinétique du $n^{ième}$ mode de \mathcal{P}_C
\mathcal{P}_{NC}	-	système non conservatif associé à la plaque
U_n^{NC}	(J)	Énergie potentielle du $n^{ième}$ mode de \mathcal{P}_{NC}
T_n^{NC}	(J)	Énergie cinétique du $n^{ième}$ mode de \mathcal{P}_{NC}
ΔU_n^{NC}	(J)	Énergie perdue par cycle par le $n^{ième}$ mode de \mathcal{P}_{NC}
Q	-	Ordre de la base polynomiale (procédure de Rayleigh-Ritz)
$p_i(x)$	-	$i^{ième}$ élément de la base polynomiale dans la direction x (procédure de Rayleigh-Ritz)
$p_j(y)$	-	$j^{ième}$ élément de la base polynomiale dans la direction y (procédure de Rayleigh-Ritz)
L_{ij}	(rad)	Coordonnées de $\Phi_x(x, y)$ dans la base polynomiale (procédure de Rayleigh-Ritz)
M_{ij}	(rad)	Coordonnées de $\Phi_y(x, y)$ dans la base polynomiale (procédure de Rayleigh-Ritz)
N_{ij}	(m)	Coordonnées de $w_0(x, y)$ dans la base polynomiale (procédure de Rayleigh-Ritz)

TAB. B.4 – Notations utilisées pour le modèle numérique de plaque

Bibliographie

- Abel, J. S. & Berners, D. P. (October 2006). A technique for nonlinear system measurement. *121th Convention of the Audio Engineering Society*, Paper 6951.
- Agganis, B. T., Muday, J. A., & Schirillo, J. A. (2010). Visual biasing of auditory localization in azimuth and depth. *Perceptual and Motor Skills*, 111(3) :872–892.
- Alais, D. & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14(3) :257–262.
- Alexandrova, I. V., Teneva, P. T., de la Rosa, S., Kloos, U., Bülthoff, H. H., & Mohler, B. J. (2010). Egocentric distance judgments in a large screen display immersive virtual environment. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, APGV '10, pages 57–60, New York, NY, USA. ACM.
- Altinsoy, M. E., Jekosch, U., Landgraf, J., & Merchel, S. (2010). Progress in auditory perception research laboratories “multimodal multimodal measurement laboratory” of dresden university of technology. In *Audio Engineering Society Convention 129*.
- Amatriain, X., Hollerer, T., Kuchera-Morin, J., & Pope, S. (2007). Immersive audio and music in the allosphere. In *International Computer Music Conference*.
- Anderson, J. S. & Bratos-Anderson, M. (2005). Radiation efficiency of rectangular orthotropic plates. *Acta Acustica United With Acustica*, 91(1) :61–76.
- Andre, J. & Rogers, S. (2006). Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Perception & Psychophysics*, 68(3) :353–361.
- Angus, J. A. S. (1999). Distributed mode loudspeakers polar patterns. *107th Convention of the Audio Engineering Society*.
- Araujo, A. L., Mota Soares, C. M., Herskovits, J., & Pedersen, P. (2009a). Visco-piezo-elastic parameter estimation in laminated plate structures. *Inverse Problems in Science and Engineering*, 17(2) :145–157.
- Araujo, A. L., Soares, C. M. M., de Freitas, M. J. M., Pedersen, P., & Herskovits, J. (2000). Combined numerical-experimental model for the identification of mechanical properties of laminated structures. *Composite Structures*, 50(4) :363–372.

- Araujo, A. L., Soares, C. M. M., & deFreitas, M. J. M. (1996). Characterization of material parameters of composite plate specimens using optimization and experimental vibrational data. *Composites Part B-engineering*, 27(2) :185–191.
- Araujo, A. L., Soares, C. M. M., Herskovits, J., & Pedersen, P. (2009b). Estimation of piezoelectric and viscoelastic properties in laminated structures. *Composite Structures*, 87(2) :168–174.
- Araujo, A. L., Soares, C. M. M., Soares, C. A. M., & Herskovits, J. (2010). Characterisation by inverse techniques of elastic, viscoelastic and piezoelectric properties of anisotropic sandwich adaptive structures. *Applied Composite Materials*, 17(5) :543–556.
- Armbruster, C., Wolter, M., Kuhlen, T., Spijkers, W., & Fimm, B. (2008). Depth perception in virtual reality : Distance estimations in peri- and extrapersonal space. *Cyberpsychology & Behavior*, 11(1) :9–15.
- Ashmead, D. H., Davis, D. L., & Northington, A. (1995). Contribution of listeners approaching motion to auditory distance perception. *Journal of Experimental Psychology-Human Perception and Performance*, 21(2) :239–256.
- Atalla, N., Nicolas, J., & Gauthier, C. (1996). Acoustic radiation of an unbaffled vibrating plate with general elastic boundary conditions. *Journal of the Acoustical Society of America*, 99(3) :1484–1494.
- Ayorinde, E. O. (1996). Elastic constants of thick orthotropic composite plates. *Journal of Composite Materials*, 30(15) :1729–1729.
- Ayorinde, E. O. & Yu, L. (2005). On the elastic characterization of composite plates with vibration data. *Journal of Sound and Vibration*, 283(1-2) :243–262.
- Backstrom, D. & Nilsson, A. C. (2007). Modelling the vibration of sandwich beams using frequency-dependent parameters. *Journal of Sound and Vibration*, 300(3-5) :589–611.
- Badeau, R., David, B., & Richard, G. (2006). A new perturbation analysis for signal enumeration in rotational invariance techniques. *IEEE Transactions on Signal Processing*, 54(2) :450–458.
- Bai, M. R. & Chung, K. C. (2002). Optimal design of panel speaker array with omnidirectional characteristics. *Journal of the Acoustical Society of America*, 112(5) :1944–1952.
- Baldis, J. J. (2001). Effects of spatial audio on memory, comprehension, and preference during desktop conferences. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '01, pages 166–173, New York, NY, USA. ACM.
- Bank, Graham ; Harris, N. (1998). The distributed mode loudspeaker-theory and practice. In *Audio Engineering Society Conference : UK 13th Conference : Microphones & Loudspeakers*.
- Barbieri, N., Barbieri, R., & Winikes, L. C. (2010). Parameters estimation of sandwich beam model with rigid polyurethane foam core. *Mechanical Systems and Signal Processing*, 24(2) :406–415.

- Barkanov, E., Skukis, E., & Petitjean, B. (2009). Characterisation of viscoelastic layers in sandwich panels via an inverse technique. *Journal of Sound and Vibration*, 327(3-5) :402–412.
- Bastos, S. F., Borges, L., & Rochinha, F. A. (2002). Numerical and experimental approach for identifying elastic parameters in sandwich plates. *Shock and Vibration*, 9(4-5) :193–201.
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A-Optics Image Science and Vision*, 20(7) :1391–1397.
- Baumgartner, S. L. & Rugh, W. J. (1975). Complete identification of a class of nonlinear-systems from steady-state frequency-response. *IEEE Transactions on Circuits and Systems*, 22(9) :753–759.
- Beall, A. C., Loomis, J. M., & Philbeck, J. W. (1994). Absolute motion parallax weakly determines visual scale. *Investigative Ophthalmology & Visual Science*, 35(4) :2111–2111.
- Bech, S., Woszczyk, W., & V., H. (1995). Interactions between audio-visual factors in a home theater system : Experimental results. *99th Convention of the Audio Engineering Society*.
- Beerends, J. G. & de Caluwe, F. E. (1999). The influence of video quality on perceived audio quality and vice versa. *Journal of the Audio Engineering Society*, 47(5) :355–362.
- Begault, D. R. (1994). *3-D sound for virtual reality and multimedia*. Academic Press Professional, Inc., San Diego, CA, USA.
- Berkhout, A. J., de Vries, D., & Vogel, P. (1993). Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America*, 93(5) :2764–2778.
- Berry, A., Guyader, J. L., & Nicolas, J. (1990). A general formulation for the sound radiation from rectangular, baffled plates with arbitrary boundary-conditions. *Journal of the Acoustical Society of America*, 88(6) :2792–2802.
- Berthaut, J., Ichchou, M. N., & Jezequel, L. (2003). Piano soundboard : structural behavior, numerical and experimental study in the modal range. *Applied Acoustics*, 64(11) :1113–1136.
- Bhat, R. B. (1985). Natural frequencies of rectangular-plates using characteristic orthogonal polynomials in rayleigh-ritz method. *Journal of Sound and Vibration*, 102(4) :493–499.
- Bicken, G., Carey, G. F., & Stearman, R. O. (2002). Frequency domain kernel estimation for 2nd-order volterra models using random multi-tone excitation. *VLSI Design*, 15(4) :701–713.
- Birman, V. (2002). On the choice of shear correction factor in sandwich structures. *Journal of Sandwich Structures & Materials*, 4(1) :83–95.
- Blauert, J. (1999). *Spatial Hearing, The Psychophysics of Human Sound Localization*. MIT Press.
- Bledzki, A. K., Kessler, A., Rikards, R., & Chate, A. (1999). Determination of elastic constants of glass/epoxy unidirectional laminates by the vibration testing of plates. *Composites Science and Technology*, 59(13) :2015–2024.

- Bohigas, O., Legrand, O., Schmit, C., & Sornette, D. (1991). Comment on spectral statistics in elastodynamics. *Journal of the Acoustical Society of America*, 89(3) :1456–1458.
- Boone, M. M. (2004). Multi-actuator panels (maps) as loudspeaker arrays for wave field synthesis. *Journal of the Audio Engineering Society*, 52(7-8) :712–723.
- Boone, Marinus M.; de Bruijn, W. P. J. (2003). Improving speech intelligibility in teleconferencing by using wave field synthesis. In *Audio Engineering Society Convention 114*.
- Bormann, K. (2005). Presence and the utility of audio spatialization. *Presence-Teleoperators and Virtual Environments*, 14(3) :278–297.
- Bormann, K. (2008). Visuals are not what they look. *Virtual Reality*, 12 :115–123. 10.1007/s10055-007-0068-4.
- Bouvier, P. (2008). The five pillars of presence : guidelins to reach presence. In *Presence*.
- Bouvier, P., Chaudeyrac, P., Loyet, R., Piranda, B., & de Sorbier, F. (2006). Immersive visual and audio world in 3d. In *9th International Conference on Computer Games*.
- Bowen, A. (2010). *Visual Localization Accuracy Determines the Bias of Auditory Targets in Azimuth and Depth*. Master's thesis, Wake Forest University.
- Bowman, D. A. & McMahan, R. P. (2007). Virtual reality : How much immersion is enough ? *Computer*, 40 :36–+.
- Boyd, S. & Chua, L. O. (1985). Fading memory and the problem of approximating nonlinear operators with volterra series. *IEEE Transactions on Circuits and Systems*, 32(11) :1150–1161.
- Boyd, S., Tang, Y. S., & Chua, L. O. (1983). Measuring volterra kernels. *IEEE Transactions on Circuits and Systems*, 30(8) :571–577.
- Bronkhorst, A. W. (January/February 2000). The cocktail party phenomenon : A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica United with Acustica*, 86 :117–128(12).
- Bronkhorst, A. W. & Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, 397(6719) :517–520.
- Brooks, F. P. (1999). What's real about virtual reality ? *IEEE Computer Graphics and Applications*, 19(6) :16–27.
- Bruneau, M. & Scelo, T. (2006). *Fundamentals of acoustics*. ISTE, London.
- Burkhardt, J. & Weaver, R. L. (1996). Spectral statistics in damped systems .2. spectral density fluctuations. *Journal of the Acoustical Society of America*, 100(1) :327–334.
- Burr, D. & Alais, D. (2006). Combining visual and auditory information. *Visual Perception, Pt 2 : Fundamentals of Awareness : Multi-Sensory Integration and High-Order Perception*, 155 :243–258.

- Burton, W. S. & Noor, A. K. (1997). Assessment of continuum models for sandwich panel honeycomb cores. *Computer Methods in Applied Mechanics and Engineering*, 145(3-4) :341–360.
- Caramiaux, B., Fdili Alaoui, S., Bouchara, T., Parseihian, G., & Rébillat, M. (2011). Gestural auditory and visual interactive platform. In *Proceedings of the 14th International Conference on Digital Audio Effects (DAFx-11), Paris, France*.
- Caron, J. (1999). *Modèle de plaques multicouches*. Cours de l'École Nationale des Ponts et Chaussées.
- Chaigne, A. & Lambourg, C. (2001). Time-domain simulation of damped impacted plates. i. theory and experiments. *Journal of the Acoustical Society of America*, 109(4) :1422–1432.
- Chatterjee, A. & Vyas, N. S. (2000). Convergence analysis of volterra series response of nonlinear systems subjected to harmonic excitation. *Journal of Sound and Vibration*, 236(2) :339–358.
- Chatterjee, A. & Vyas, N. S. (2003). Non-linear parameter estimation with volterra series using the method of recursive iteration through harmonic probing. *Journal of Sound and Vibration*, 268(4) :657–678.
- Chen, H. W. (1995). Modeling and identification of parallel nonlinear-systems - structural classification and parameter-estimation methods. *Proceedings of the IEEE*, 83(1) :39–66.
- Chen, S., Billings, S. A., & Luo, W. (1989). Orthogonal least-squares methods and their application to non-linear system-identification. *International Journal of Control*, 50(5) :1873–1896.
- Colloms, M., Panzer, J., Gotcharov, V., & Taylor, V. (1998). Distortion mechanisms of distributed mode (dm) panel loudspeakers. *104th Convention of the Audio Engineering Society*.
- Conant, A., Chu, W., Silva, R., McNally, T., Sparks, G., & Hoover, K. (2008). Acoustics of a large immersive environment : the allosphere. In *155th meeting of the Acoustical Society of America*.
- Corteel, E. (2004). *Caractérisation et Extensions de la Wave Field Synthesis en conditions réelles*. PhD thesis, Université de Paris 6.
- Corteel, E. (2006). Equalization in an extended area using multichannel inversion and wave field synthesis. *Journal of the Audio Engineering Society*, 54(12) :1140–1161.
- Corteel, E., Horbach, U., & Pellegrini, R. (2002). Multichannel inverse filtering of multiexciter distributed mode loudspeakers for wave field synthesis. *112th Convention of the Audio Engineering Society*.
- Corteel, E., Kuhn-Rahloff, C., & Pellegrini, R. (2008). Wave field synthesis rendering with increased aliasing frequency. *124th Convention of the Audio Engineering Society*.
- Corteel, E., NGuyen, K. V., Warusfel, O., Caulkins, T., & Pellegrini, R. (2007). Objective and subjective comparison of electrodynamic and map loudspeakers for wave field synthesis. *30th International Conference of the Audio Engineering Society*.
- Cortes, F. & Elejabarrieta, M. J. (2006). An approximate numerical method for the complex eigenproblem in systems characterised by a structural damping matrix. *Journal of Sound and Vibration*, 296(1-2) :166–182.

- Courgeon, M., Martin, J.-C., & Jacquemin, C. (2008). MARC : a Multimodal Affective and Reactive Character. In *Proceedings of the 1st Workshop on AFFective Interaction in Natural Environnements (AFFINE)*.
- Courgeon, M., Rébillat, M., Katz, B. F., Clavel, C., & Martin, J.-C. (2010). Life-sized audiovisual spatial social scenes with multiple characters : MARC & SMART-I². In *5^{èmes} journées de l'Association Française de Réalité Virtuelle (AFRV)*.
- Craven, P. G. & Gerzon, M. A. (September 1992). Practical adaptive room and loudspeaker equaliser for HI-FI use. *7th Conference of the Audio Engineering Society : Digital Signal Processing*.
- Creem-Regehr, S. H., Willemsen, P., Gooch, A. A., & Thompson, W. B. (2005). The influence of restricted viewing conditions on egocentric distance perception : Implications for real and virtual indoor environments. *Perception*, 34(2) :191–204.
- Cremer, L., Heckl, M., & Peterson, B. (2005). *Structure-Borne Sound : Structural Vibrations and Sound Radiation at Audio Frequencies Third Edition*. Springer-Verlag.
- Crighton, D. G. (1989). The 1988 rayleigh medal lecture - fluid loading - the interaction between sound and vibration. *Journal of Sound and Vibration*, 133(1) :1–27.
- Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). Surround-screen projection-based virtual reality : the design and implementation of the cave. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '93, pages 135–142, New York, NY, USA. ACM.
- Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The cave : audio visual experience automatic virtual environment. *Commun. ACM*, 35 :64–72.
- Côté, N., Koehl, V., Paquier, M., & Devillers, F. (2011). Interaction between auditory and visual distance cues in virtual reality applications. In *Proceedings of the Forum Acusticum, Aalborg, Danemark*.
- Cugnoni, J., Gmur, T., & Schorderet, A. (2004). Identification by modal analysis of composite structures modelled with fsdt and hsdt laminated shell finite elements. *Composites Part A-applied Science and Manufacturing*, 35(7-8) :977–987.
- Cugnoni, J., Gmur, T., & Schorderet, A. (2007). Inverse method based on modal analysis for characterizing the constitutive properties of thick composite plates. *Computers & Structures*, 85(17-18) :1310–1320.
- Cunha, J. & Piranda, J. (1999). Application of model updating techniques in dynamics for the identification of elastic constants of composite materials. *Composites Part B-engineering*, 30(1) :79–85.
- Cutting, J. E. (1997). How the eye measures reality and virtual reality. *Behavior Research Methods Instruments & Computers*, 29(1) :27–36.
- Czerwinski, E., Voishvillo, A., Alexandrov, S., & Terekhov, A. (2001). Multitone testing of sound system components some results and conclusions - part 2 : Modeling and application. *Journal of the Audio Engineering Society*, 49(12) :1181–1192.

- Daniel, I. M. & Abot, J. L. (2000). Fabrication, testing and analysis of composite sandwich beams. *Composites Science and Technology*, 60(12-13) :2455–2463.
- Daniel, J. (2001). *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD thesis, Université de Paris 6.
- Darken, R. P., Bernatovitch, D., Lawson, J. P., & Peterson, B. (1999). Quantitative measures of presence in virtual environments : The roles of attention and spatial comprehension. *CyberPsychology & Behavior*, 2(4) :337–347. PMID : 19178230.
- Darken, R. P. & Sibert, J. L. (1993). A toolset for navigation in virtual environments. In *Proceedings of the 6th annual ACM symposium on User interface software and technology*, UIST '93, pages 157–165, New York, NY, USA. ACM.
- de Bruijn, W., Boone, M. M., & de Vries D. (2000). Sound localization in a videoconferencing system based on wave field synthesis. *112th Convention of the Audio Engineering Society*.
- de Bruijn, W. P. J. & Boone, M. M. (2002). Subjective experiments on the effects of combining spatialized audio and 2d video projection in audio-visual systems. *112th Convention of the Audio Engineering Society*.
- de Bruijn, W. P. J. & Boone, M. M. (2003). Application of wave field synthesis in life-size videoconferencing. *114th Convention of the Audio Engineering Society*.
- de Sorbier de Pougnaudresse, F., Bouvier, P., Herubel, A., & Biri, V. (2009). From research on the virtual reality installation. In Hippe, Z. & Kulikowski, J., editors, *Human-Computer Systems Interaction*, volume 60 of *Advances in Soft Computing*, pages 335–345. Springer Berlin / Heidelberg.
- de Visscher, J., Sol, H., de Wilde, W. P., & Vantomme, J. (1997). Identification of the damping properties of orthotropic composite materials using a mixed numerical experimental method. *Applied Composite Materials*, 4(1) :13–33.
- de Vries, D. (2009). *Wave Field Synthesis*. Audio Engineering society.
- Deobald, L. R. & Gibson, R. F. (1988). Determination of elastic-constants of orthotropic plates by a modal-analysis rayleigh-ritz technique. *Journal of Sound and Vibration*, 124(2) :269–283.
- Dinh, H. Q., Walker, N., Song, C., Kobayashi, A., & Hodges, L. F. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. *Virtual Reality Conference, IEEE*, 0 :222.
- dos Santos, J. V. A., Araujo, A. L., & Soares, C. M. M. (1998). Eigenfrequency analysis of completely free multilayered rectangular plates using a higher-order model and ritz technique. *Mechanics of Composite Materials and Structures*, 5(1) :55–80.
- Ege, K., Boutillon, X., & David, B. (2009). High-resolution modal analysis. *Journal of Sound and Vibration*, 325(4-5) :852–869.

- Ege, K., Boutillon, X., & Rébillat, M. (2011). Vibroacoustics of the piano soundboard (part 1) : non-linearity and modal properties in the low- and mid-frequency ranges. *En préparation pour le Journal of Sound and Vibration*.
- Evrard, M., André, C. R., Verly, J. G., & J.-J., E. (2011). Object-based sound re-mix for spatially coherent audio rendering of an existing stereoscopic-3-d animation movie. In *131st Convention of the Audio Engineering Society*.
- Faria, R. R. A., Thomaz, L. F., Soares, L., Santos, B. T., Zuffo, M. K., & Zuffo, J. A. (2005a). Audience - audio immersion experiences in the caverna digital. *Proceedings of the 10th Brazilian Symposium on Computer Music, Belo Horizonte*, pages 106–117.
- Faria, R. R. A., Zuffo, M. K., & Zuffo, J. A. (2005b). Improving spatial perception through sound field simulation in vr. *Proceedings of the 2005 IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, pages 103–108.
- Farina, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. *108th Convention of the Audio Engineering Society*.
- Farina, A., Bellini, A., & Armelloni, E. (2001). Nonlinear convolution : a new approach for the auralization of distorting systems. *110th Convention of the Audio Engineering Society*.
- Frederiksen, P. S. (1997a). Application of an improved model for the identification of material parameters. *Mechanics of Composite Materials and Structures*, 4(4) :297–316.
- Frederiksen, P. S. (1997b). Experimental procedure and results for the identification of elastic constants of thick orthotropic plates. *Journal of Composite Materials*, 31(4) :360–382.
- Frederiksen, P. S. (1997c). Numerical studies for the identification of orthotropic elastic constants of thick plates. *European Journal of Mechanics A-Solids*, 16(1) :117–140.
- Frederiksen, P. S. (1998). Parameter uncertainty and design of optimal experiments for the estimation of elastic constants. *International Journal of Solids and Structures*, 35(12) :1241–1260.
- Freeman, J. & Lessiter, J. (2001). Here, there and everywhere : the effects of multi-channel audio on presence. In *Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29-August 1, 2001*.
- Fuchs, P. (2006). *Traité de la réalité virtuelle*, volume 2, chapter La restitution visuelle stéréoscopique. Presses de l'École des Mines de Paris, Laval, France, 3 edition.
- Fuchs, P., Benali-Koudja, M., Hafez, M., Papin, J.-P., Tsingos, N., & Warusfel, O. (2006a). *Traité de la réalité virtuelle*, volume 1, chapter Les sens de l'homme. Presses de l'École des Mines de Paris, Laval, France, 3 edition.
- Fuchs, P., Berthoz, A., & Verchez, J. (2006b). *Traité de la réalité virtuelle*, volume 1, chapter Introduction à la réalité virtuelle. Presses de l'École des Mines de Paris, Laval, France, 3 edition.

- Fuchs, P. & Dominjon, L. (2006). *Traité de la réalité virtuelle*, volume 4, chapter Les interfaces visuelles. Presses de l'École des Mines de Paris, Laval, France, 3 edition.
- Fukushima, S. S., Loomis, J. M., & DaSilva, J. A. (1997). Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology-Human Perception and Performance*, 23(1) :86–100.
- Gagneja, S., Gibson, R. F., & Ayorinde, E. O. (2001). Design of test specimens for the determination of elastic through-thickness shear properties of thick composites from measured modal vibration frequencies. *Composites Science and Technology*, 61(5) :679–687.
- Gallman, P. G. (1975). Iterative method for identification of nonlinear-systems using a uryson model. *IEEE Transactions on Automatic Control*, 20(6) :771–775.
- Gardner, M. B. (1968). Proximity image effect in sound localization. *Journal of the Acoustical Society of America*, 43(1) :163–&.
- Gee, K. L., Sparrow, V. W., James, M. M., Downing, J. M., & Hobbs, C. M. (2006). Measurement and prediction of nonlinearity in outdoor propagation of periodic signals. *Journal of the Acoustical Society of America*, 120(5) :2491–2499.
- Geradin, M. & Rixen, D. (1997). *Mechanical Vibrations : Theory and Applications to Structural Dynamics*.
- Gerzon, M. A. (1985). Ambisonics in multichannel broadcasting and video. *Journal of the Audio Engineering Socitey*, 33(11) :859–871.
- Gibson, L. J. & Ashby, M. (2001). *Cellular Solids : Structure and Properties, second edition*. Cambridge University Press.
- Gilbert, J., Menguy, L., & Campbell, M. (2008). A simulation tool for brassiness studies (I). *Journal of the Acoustical Society of America*, 123(4) :1854–1857.
- Gilkey, R. H. & Weisenberger, J. M. (1995). The sense of presence for the suddenly deafened adult - implications for virtual environments. *Presence-Teleoperators and Virtual Environments*, 4(4) :357–363.
- Giraudeau, A., Pierron, F., & Guo, B. Q. (2010). An alternative to modal analysis for material stiffness and damping identification from vibrating plates. *Journal of Sound and Vibration*, 329(10) :1653–1672.
- Godfroy, M., Roumes, C., & Dauchy, P. (2003). Spatial variations of visual-auditory fusion areas. *Perception*, 32(10) :1233–1245.
- Godfroy, M., Welch, R., Sandor, P., & Roumes, C. (2011). The interaction of vision and audition in two-dimensional space. *Draft en cours de publication*.
- Gogel, W. C. (1969). Sensing of retinal size. *Vision Research*, 9(9) :1079–1088.
- Golub, G. (1965-06-29). Numerical methods for solving linear least squares problems.

- Grechkin, T. Y., Nguyen, T. D., Plumert, J. M., Cremer, J. F., & Kearney, J. K. (2010). How does presentation method and measurement protocol affect distance estimation in real and virtual environments ? *ACM Transactions on Applied Perception*, 7(4) :26.
- Grediac, M., Fournier, N., Paris, P. A., & Surrrel, Y. (1998). Direct identification of elastic constants of anisotropic plates by modal analysis : Experimental results. *Journal of Sound and Vibration*, 210(5) :643–659.
- Griesinger, D. (1996). Beyond mls - occupied hall measurement with fft techniques. *101th Convention of the Audio Engineering Society*.
- Gröhn, M. (2006). *Application of spatial sound reproduction in virtual environments - experiments in localization, navigation, and orientation*. PhD thesis, Department of Computer Science and Engineering, Helsinki University of Technology, Espoo, Finland.
- Gröhn, M., Lokki, T., & Takala, T. (2005). Comparison of auditory, visual, and audiovisual navigation in a 3d space. *ACM Transactions on Applied Perception*, 2 :564–570.
- Guastavino, C. & Katz, B. F. G. (2004). Perceptual evaluation of multi-dimensional spatial audio reproduction. *Journal of the Acoustical Society of America*, 116(2) :1105–1115.
- Guastavino, C., Katz, B. F. G., Polack, J. D., Levitin, D. J., & Dubois, D. (2005). Ecological validity of soundscape reproduction. *Acta Acustica United with Acustica*, 91(2) :333–341.
- Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*, 15(1) :20–29.
- Harris, N. J. & Hawksford, M. O. J. (2000). Introduction to distributed mode loudspeakers (dml) with first-order behavioural modelling. *IEE Proceedings-circuits Devices and Systems*, 147(3) :153–157.
- Hartnagel, D. (2007). *La perception de l'espace multisensoriel appréhendée par l'étude de la fusion visuo-auditive : les effets de la dissociation des référentiels spatiaux*. PhD thesis, Université PARIS 8 - Saint-Denis, Vincennes.
- Hasler, M. (1999). *Phénomènes non linéaires, Chapitre 3 : Séries de Volterra*. EPFL Lausanne.
- Hause, T. & Librescu, L. (2006). Flexural free vibration of sandwich flat panels with laminated anisotropic face sheets. *Journal of Sound and Vibration*, 297(3-5) :823–841.
- Hawksford, M. O. J. (2005). System measurement and identification using pseudorandom filtered noise and music sequences. *Journal of the Audio Engineering Society*, 53(4) :275–296.
- Hawley, M. L., Litovsky, R. Y., & Colburn, H. S. (1999). Speech intelligibility and localization in a multi-source environment. *Journal of the Acoustical Society of America*, 105(6) :3436–3448.
- Hendrix, C. & Barfield, W. (1995). Presence in virtual environments as a function of visual and auditory cues. *Virtual Reality Annual International Symposium*, 0 :74.

- Hendrix, C. & Barfield, W. (1996). The sense of presence within auditory virtual environments. *Presence-Teleoperators and Virtual Environments*, 5(3) :290–301.
- Heron, J., Whitaker, D., McGraw, P. V., & Horoshenkov, K. V. (2007). Adaptation minimizes distance-related audiovisual delays. *Journal of Vision*, 7(13) :5.
- Hiipakka, J., Ilmonen, T., Lokki, T., Gröhn, M., & Savioja, L. (2001). Implementation issues of 3d audio in a virtual room. In A. J. Woods, M. T. Bolas, J. O. Merritt, & S. A. Benton, editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 4297 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 486–495.
- Hiipakka, J., Ilmonen, T., Lokki, T., & Savioja, L. (2000). Sound signal processing for a virtual room. In *EUPSICO 2000 : European signal processing conference No 10, Tampere , FINLANDE*.
- Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3) :33.
- Hollier, M. P., Rimell, A. N., Hands, D. S., & Voelcker, R. M. (1999). Multi-modal perception. *BT Technology Journal*, 17(1) :35–46.
- Horbach, U., Corteel, E., & de Vries, D. (2002). Spatial audio reproduction using distributed mode loudspeaker arrays. *21th conference of the Audio Engineering Society*.
- Howarth, P. A. (2011). Potential hazards of viewing 3-d stereoscopic television, cinema and computer games : a review. *Ophthalmic and Physiological Optics*, 31(2) :111–122.
- Hwang, S. F. & Chang, C. S. (2000). Determination of elastic constants of materials by vibration testing. *Composite Structures*, 49(2) :183–190.
- Ichchou, M. N., Bareille, O., & Berthaut, J. (2008). Identification of effective sandwich structural properties via an inverse wave approach. *Engineering Structures*, 30(10) :2591–2604.
- IJsselstein, W. A., de Ridder, H., Freeman, J., & Avons, S. E. (2000). Presence : Concept, determinants and measurement. *Human Vision and Electronic Imaging V*, 3959 :520–529.
- Insko, E. B. (2003). *Being There : Concepts, effects and measurement of user presence in synthetic environments*. IOS Press, Amsterdam.
- Interrante, V., Ries, B., Lindquist, J., Kaeding, M., & Anderson, L. (2008). Elucidating factors that can facilitate veridical spatial perception in immersive virtual environments. *Presence-Teleoperators and Virtual Environments*, 17(2) :176–198.
- Jack, C. E. & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on ventriloquism effect. *Perceptual and Motor Skills*, 37(3) :967–979.
- Johnson, C. D. & Kienholz, D. A. (1982). Finite-element prediction of damping in structures with constrained viscoelastic layers. *AIAA Journal*, 20(9) :1284–1290.

- Jot, J.-M. (1999). Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. *Multimedia Systems*, 7 :55–69. 10.1007/s005300050111.
- Junger, M. C. & Feit, D. (1986). *Sound, Structures, and Their Interaction*.
- Kaizer, A. J. M. (1987). Modeling of the nonlinear response of an electrodynamic loudspeaker by a volterra series expansion. *Journal of the Audio Engineering Society*, 35(6) :421–433.
- Kayran, A. H. & Eksioglu, E. M. (2005). Nonlinear system identification using deterministic multilevel sequences. *Circuits Systems and Signal Processing*, 24(2) :151–181.
- Kearney, G., Gorzel, M., Boland, F., & Rice, H. (2010). Depth perception in interactive virtual acoustic environments using higher order ambisonic soundfields. In *Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics*.
- Kelley, C. (1999). *Iterative Methods for Optimization*. Society for Industrial and Applied Mathematics, Philadelphia.
- Kilgore, R., Chignell, M., & Smith, P. (2003). Spatialized audioconferencing : what are the benefits ? In *Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research*, CASCON '03, pages 135–144. IBM Press.
- Kilgore, R. M. (2009). Simple displays of talker location improve voice identification performance in multitalker, spatialized audio environments. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 51(2) :224–239.
- Kim, C. S. (1986). Natural frequencies of rectangular-plates using characteristic orthogonal polynomials in rayleigh-ritz method - comment. *Journal of Sound and Vibration*, 108(1) :166–168.
- Kim, H. Y., Suzuki, Y., Takane, S., & Sone, T. (2001). Control of auditory distance perception based on the auditory parallax model. *Applied Acoustics*, 62(3) :245–270.
- Kim, J. S. (2007). Free vibration of laminated and sandwich plates using enhanced plate theories. *Journal of Sound and Vibration*, 308(1-2) :268–286.
- Klein, E., Swan, J. E., Schmidt, G. S., Livingston, M. A., & Staadt, O. G. (2009). Measurement protocols for medium-field distance perception in large-screen immersive displays. *IEEE Virtual Reality 2009, Proceedings*, pages 107–113.
- Klippel, W. (1996a). Modeling the nonlinearities in horn loudspeakers. *Journal of the Audio Engineering Society*, 44(6) :470–480.
- Klippel, W. (1996b). Nonlinear system identification for horn loudspeakers. *Journal of the Audio Engineering Society*, 44(10) :811–820.
- Klippel, W. (2006). Tutorial : Loudspeaker nonlinearities - causes, parameters, symptoms. *Journal of the Audio Engineering Society*, 54(10) :907–939.

- Kohlrausch, A. & van de Par, S. (2005). *Communication Acoustics (Signals and Communication Technology), Chapter 5*. Springer-Verlag.
- Komiyama, S. (1989). Subjective evaluation of angular displacement between picture and sound directions for hdtv sound systems. *Journal of the Audio Engineering Society*, 37(4) :210–214.
- Komiyama, S., Morita, A., Kurozumi, K., & Nakabayashi, K. (1991). Distance control system for a sound image. In *Audio Engineering Society Conference : 9th International Conference : Television Sound Today and Tomorrow*.
- Kopinska, A. & Harris, L. R. (2004). Simultaneity constancy. *Perception*, 33(9) :1049–1060.
- Kuhlen, T., Assenmacher, I., & Lentz, T. (2007). A true spatial sound system for cave-like displays using four loudspeakers. *Virtual Reality, Proceedings*, 4563 :270–279.
- Kunz, B. R., Wouters, L., Smith, D., Thompson, W. B., & Creem-Regehr, S. H. (2009). Revisiting the effect of quality of graphics on distance judgments in virtual environments : A comparison of verbal reports and blind walking. *Attention Perception & Psychophysics*, 71(6) :1284–1293.
- Kuster, M., De Vries, D., Beer, D., & Brix, S. (2006). Structural and acoustic analysis of multiactuator panels. *Journal of the Audio Engineering Society*, 54(11) :1065–1076.
- Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9 :112–147.
- Lai, T. C. & Ip, K. H. (1996). Parameter estimation of orthotropic plates by bayesian sensitivity analysis. *Composite Structures*, 34(1) :29–42.
- Laroche, J. (1993). The use of the matrix pencil method for the spectrum analysis of musical signals. *Journal of the Acoustical Society of America*, 94(4) :1958–1965.
- Larsson, Pontus ; Västfjäll, D. K. M. (2002). Better presence and performance in virtual environments by improved binaural sound rendering. In *Audio Engineering Society Conference : 22nd International Conference : Virtual, Synthetic, and Entertainment Audio*.
- Larsson, P., Västfjäll, D., Olsson, P., & Kleiner, M. (2007). When what you hear is what you see : Presence and auditory-visual integration in virtual environments. In *Proceedings of the 10th Annual International Workshop on Presence*.
- Larsson, P., Väjamäe, A., Västfjäll, D., & Kleiner, M. (2005). *Immersed in Media Experiences : Presence Psychology and Design (Handbook of Presence)*, chapter Auditory-induced presence in mediated environments and related technology. Lawrence Erlbaum Associates, Mahwah.
- Larsson, P., Västfjäll, D., & Kleiner, M. (2001). Ecological acoustics and the multi-modal perception of rooms : real and unreal experiences of auditory-visual virtual environments. In *Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29-August 1, 2001*.

- Laulagnet, B. (1998). Sound radiation by a simply supported unbaffled plate. *Journal of the Acoustical Society of America*, 103(5) :2451–2462.
- Lauwagie, T., Sol, H., & Heylen, W. (2006). Handling uncertainties in mixed numerical-experimental techniques for vibration based material identification. *Journal of Sound and Vibration*, 291(3-5) :723–739.
- Lentz, T., Schroder, D., Vorlander, M., & Assenmacher, I. (2007). Virtual reality system with integrated sound field simulation and reproduction. *EURASIP Journal on Advances in Signal Processing*, page 70540.
- Lessiter, J. & Freeman, J. (2001). Really hear ? the effects of audio quality on presence. In *Proceedings of the Fourth Annual International Workshop on Presence*, pages 288–324.
- Lesueur, C. (1988). *Rayonnement acoustique des structures*.
- Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioural Brain Research*, 121(1-2) :69–79.
- Lewald, J. & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, 16(3) :468–478.
- Lewald, J. & Guski, R. (2004). Auditory-visual temporal integration as a function of distance : no compensation for sound-transmission time in human perception. *Neuroscience Letters*, 357(2) :119–122.
- Liu, Q. L. & Zhao, Y. (2006). Natural frequency analysis of a sandwich panel with soft core based on a refined shear deformation model. *Composite Structures*, 72(3) :364–374.
- Lokki, T., Gröhn, M., Savioja, L., & Takala, T. (2000). A case study of auditory navigation in virtual acoustic environments. *International Conference on Auditory Display*.
- Lokki, T. & Grohn, M. (2005). Navigation with auditory cues in a virtual environment. *IEEE Multimedia*, 12(2) :80–86.
- Lombard, M. & Ditton, T. (1997). At the heart of it all : The concept of presence. *Journal of Computer-Mediated Communication*, 3(2) :0–0.
- Loomis, J. M., Klatzky, R. L., Philbeck, J. W., & Golledge, R. G. (1998). Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics*, 60(6) :966–980.
- Loomis, J. M. & Knapp, J. M. (2003). *Virtual and Adaptive Environments : Applications, Implications, and Human Performance Issues*. Lawrence Erlbaum Associates.
- Lopez, J., Pueo, B., & Cobos, M. (September 2008). Conventional and distributed mode loudspeaker arrays for the application of wave-field synthesis to videoconference. *Proceedings Of IWAENC*.
- Lopez, J. J., Cobos, M., & Pueo, B. (2010). Elevation in wave-field synthesis using hrtf cues. *Acta Acustica united with Acustica*, 96(2) :340–350.

- Ma, F., Imam, A., & Morzfeld, M. (2009). The decoupling of damped linear systems in oscillatory free vibration. *Journal of Sound and Vibration*, 324(1-2) :408–428.
- Mac Intire, M. E. & Woodhouse, J. (1988). On measuring the elastic and damping constants of orthotropic sheet materials. *Acta Metallurgica*, 36(6) :1397–1416.
- MacDonald, J. A., Balakrishnan, J. D., Orosz, M. D., & Karplus, W. J. (2002). Intelligibility of speech in a virtual 3-d environment. *Human Factors*, 44(2) :272–286.
- Majdak, P., Balazs, P., & Laback, B. (2007). Multiple exponential sweep method for fast measurement of head-related transfer functions. *Journal of the Audio Engineering Society*, 55(7-8) :623–637.
- Mattei, P. O. (2007). *Cours de Vibroacoustique*.
- Matter, M., Gmur, T., Cugnoni, J., & Schorderet, A. (2009). Numerical-experimental identification of the elastic and damping properties in composite plates. *Composite Structures*, 90(2) :180–187.
- Matter, M., Gmur, T., Cugnoni, J., & Schorderet, A. (2011). Identification of the elastic and damping properties in sandwich structures with a low core-to-skin stiffness ratio. *Composite Structures*, 93(2) :331–341.
- Mead, D. J. (1982). A comparison of some equations for the flexural vibration of damped sandwich beams. *Journal of Sound and Vibration*, 83(3) :363–377.
- Mead, D. J. & Markus, S. (1969). Forced vibration of a 3-layer, damped sandwich beam with arbitrary boundary conditions. *Journal of Sound and Vibration*, 10(2) :163–&.
- Melchior, F., Brix, S., Sporer, T., Röder, T., & Klehs, B. (2003). Wave field synthesis in combination with 2d video projection. *24th International Conference of the Audio Engineering Society*.
- Mershon, D. H., Desaulniers, D., Amerson, T. L., & Kiefer, S. A. (1980). Visual capture in auditory distance perception : proximity image effect reconsidered. *Journal of Auditory Research*, 20(2) :129–136.
- Moller, H. (1989). Reproduction of artificial-head recordings through loudspeakers. *Journal of the Audio Engineering Society*, 37(1/2) :30–33.
- Morise, M., Irino, T., Banno, H., & Kawahara, H. (2007). Warped-tsp : An acoustic measurement signal robust to background noise and harmonic distortion. *Electronics and Communications in Japan Part III-Fundamental Electronic Science*, 90(4) :18–26.
- Moussu, F. & Nivoit, M. (1993). Determination of elastic-constants of orthotropic plates by a modal-analysis method of superposition. *Journal of Sound and Vibration*, 165(1) :149–163.
- Muller, S. & Massarani, P. (2001). Transfer-function measurement with sweeps. *Journal of the Audio Engineering Society*, 49(6) :443–471.
- Muravskii, G. B. (2004). On frequency independent damping. *Journal of Sound and Vibration*, 274(3-5) :653–668.

- Murray, C. D., Arnold, P., & Thornton, B. (2000). Presence accompanying induced hearing loss : Implications for immersive virtual environments. *Presence-Teleoperators and Virtual Environments*, 9(2) :137–148.
- Muthurajan, K. G., Sanakaranarayananasamy, K., & Rao, B. N. (2004). Evaluation of elastic constants of specially orthotropic plates through vibration testing. *Journal of Sound and Vibration*, 272(1-2) :413–424.
- Naceri, A., Chellali, R., Dionnet, F., & Toma, S. (2009). Depth perception within virtual environments : a comparative study between wide screen stereoscopic displays and head mounted devices. *2009 Computation World : Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns*, pages 460–466.
- Naef, M., Staadt, O., & Gross, M. (2002). Spatialized audio rendering for immersive virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology*, VRST '02, pages 65–72, New York, NY, USA. ACM.
- Nawrot, M. & Stroyan, K. (2009). The motion/pursuit law for visual depth perception from motion parallax. *Vision Research*, 49(15) :1969–1978.
- Nelisse, H., Beslin, O., & Nicolas, J. (1998). A generalized approach for the acoustic radiation from a baffled or unbaffled plate with arbitrary boundary conditions, immersed in a light or heavy fluid. *Journal of Sound and Vibration*, 211(2) :207–225.
- Nguyen, K. (2008). Integrating visual and auditory spatial cues in a virtual reality environment. *Journal of Virtual Reality and Broadcasting*.
- Nicol, R. (2010). *Binaural technology*. Audio Engineering society.
- Nilsson, A. C. (1990). Wave-propagation in and sound-transmission through sandwich plates. *Journal of Sound and Vibration*, 138(1) :73–94.
- Nilsson, E. & Nilsson, A. C. (2002). Prediction and measurement of some dynamic properties of sandwich structures with honeycomb and foam cores. *Journal of Sound and Vibration*, 251(3) :409–430.
- Noguès, M., Corteel, E., & Warusfel, O. (2003). Monitoring distance effect with wave field synthesis. In *6th Int. Conference on Digital Audio Effects (DAFX-03)*, London, UK.
- Norcross, S. & Vanderkooy, J. (1995). A survey on the effects of nonlinearity on various types of transfer-function measurements. *99th Convention of the Audio Engineering Society*.
- Novák, A. (2009). *Identification of Nonlinear Systems in Acoustics*. PhD thesis, Université du Maine, Le Mans, France.
- Novák, A., Simon, L., Kadlec, F., & Lotton, P. (2010a). Nonlinear system identification using exponential swept-sine signal. *IEEE Transactions on Instrumentation and Measurement*, 59(8) :2220–2229.

- Novák, A., Simon, L., & Lotton, P. (2010b). Analysis, synthesis, and classification of nonlinear systems using synchronized swept-sine method for audio effects. *EURASP Journal on Advances in Signal Processing*, page 793816.
- Novák, A., Simon, L., Lotton, P., & F., K. (2009). Modeling of nonlinear audio systems using swept-sine signals : application to audio effects. *12th International Conference on Audio Effects*.
- Ogi, T., Kayahara, T., Kato, M., Asayama, H., & Hirose, M. (2003). Immersive sound field simulation in multi-screen projection displays. In *Proceedings of the workshop on Virtual environments 2003*, EGVE '03, pages 135–142, New York, NY, USA. ACM.
- Okamoto, T., Cui, Z. L., Iwaya, Y., & Suzuki, Y. (2010). Implementation of a high-definition 3d audio-visual display based on higher-order ambisonics using a 157-loudspeaker array combined with a 3d projection display. In *Network Infrastructure and Digital Content, 2010 2nd IEEE International Conference on*, pages 179 –183.
- Okubo, H., Nakayama, Y., Naito, Y., Ikenaga, T., & Komiyama, S. (2005). An application of lined-up loudspeaker array system fo mixed reality audio-visual reproduction system. *119th Convention of the Audio Engineering Society*.
- Ozawa, K., Chujo, Y., Suzuki, Y., & Sone, T. (2003). Psychological factors involved in auditory presence. *Acoustical Science and Technology*, 24(1) :42–44.
- Ozawa, K. & Miyasaka, M. (2004). Effects of reproduced sound pressure levels on auditory presence. *Acoustical Science and Technology*, 25(3) :207–209.
- Pagnacco, E., Moreau, A., & Lemosse, D. (2007). Inverse strategies for the identification of elastic and viscoelastic material parameters using full-field measurements. *Materials Science and Engineering A-structural Materials Properties Microstructure and Processing*, 452 :737–745.
- Palm, G. (1978). On representation and approximation of nonlinear systems. *Biological Cybernetics*, 31 :119–124.
- Palm, G. (1979). On representation and approximation of nonlinear systems, part ii : Discrete time. *Biological Cybernetics*, 34 :49–52.
- Pellegrini, S. R. (2001). Quality assessment of auditory virtual environments. *International Conference on Auditory Display*.
- Perrott, D. R. (1993). Auditory and visual localization : Two modalities, one world. *12th International Conference of the Audio Engineering Society*.
- Phani, A. S. & Adhikari, S. (2008). Rayleigh quotient and dissipative systems. *Journal of Applied Mechanics*, 75(6) :061005.
- Plumert, J. M., Kearney, J. K., Cremer, J. F., & Recker, K. (2005). Distance perception in real and virtual environments. *ACM Transactions on Applied Perception*, 2 :216–233.

- Porschmann, C. & Storig, C. (2009). Investigations into the velocity and distance perception of moving sound sources. *Acta Acustica United with Acustica*, 95(4) :696–706.
- Prazenica, R. J. & Kurdila, A. J. (2004). Volterra kernel identification using triangular wavelets. *Journal of Vibration and Control*, 10(4) :597–622.
- Pritz, T. (1998). Frequency dependences of complex moduli and complex poisson's ratio of real solid materials. *Journal of Sound and Vibration*, 214(1) :83–104.
- Pueo, B., Lopez, J. J., & Escolano, J. (2008). Edge boundary conditions impact on the radiation of multiactuator panels for multichannel audio reproduction. *Acta Acustica United with Acustica*, 94(5) :754–764.
- Pueo, B., Lopez, J. J., Escolano, J., & Bleda, S. (2007). Analysis of multiactuator panels in the space-time wavenumber domain. *Journal of the Audio Engineering Society*, 55(12) :1092–1106.
- Pueo, B., Lopez, J. J., Escolano, J., & Horchens, L. (2010a). Multiactuator panels for wave field synthesis : Evolution and present developments. *Journal of the Audio Engineering Society*, 58(12) :1045–1063.
- Pueo, B., Lopez, J. J., Ramos, G., & Escolano, J. (2009). Efficient equalization of multi-exciter distributed mode loudspeakers. *Applied Acoustics*, 70(5) :737–746.
- Pueo, B., Ramos, G., & Lopez, J. J. (2010b). Strategies for bass enhancement in multiactuator panels for wave field synthesis. *Applied Acoustics*, 71(5) :722–730.
- Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6) :456–466.
- Putra, A. & Thompson, D. J. (2010). Sound radiation from rectangular baffled and unbaffled plates. *Applied Acoustics*, 71(12) :1113–1125.
- Qian, G. L., Hoa, S. V., & Xiao, X. R. (1997). A vibration method for measuring mechanical properties of composite, theory and experiment. *Composite Structures*, 39(1-2) :31–38.
- Quaegebeur, N. & Chaigne, A. (2008). Nonlinear vibrations of loudspeaker-like structures. *Journal of Sound and Vibration*, 309(1-2) :178–196.
- Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Cahiers De Psychologie Cognitive-Current Psychology of Cognition*, 13(1) :3–51.
- Rébillat, M. & Boutillon, X. (2010a). Identification of elastic and damping properties of sandwich structures based on high resolution modal analysis of point measurements. In *ISMA, Leuven, Belgium*.
- Rébillat, M. & Boutillon, X. (2010b). Identification of honeycomb sandwich properties by high-resolution modal analysis. In *ICEM, Poitiers, France*.
- Rébillat, M. & Boutillon, X. (2011a). Measurement of relevant elastic and damping material properties in sandwich thick-plates. *Journal of Sound and Vibration*, 330(25) :6098–6121.

- Rébillat, M. & Boutillon, X. (2011b). A method to measure elastic and dissipative material properties of sandwich structures and its numerical validation. In *CSMA, Giens, France*.
- Rébillat, M., Boutillon, X., Corteel, E., & Katz, B. F. G. (2011a). Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments. *Soumis à ACM Transactions on Applied Perception*.
- Rébillat, M., Boutillon, X., Corteel, E., & Katz, B. F. G. (2011b). Audio, visual and audio-visual egocentric distance perception in virtual environments. In *Proceedings of the Forum Acusticum, Aalborg, Denmark*.
- Rébillat, M., Corteel, E., & Katz, B. (2008). SMART-I² : “spatial multi-user audio-visual real-time interactive interface”. *125th Convention of the Audio Engineering Society*.
- Rébillat, M., Corteel, E., & Katz, B. (2010a). The SMART-I² : A new approach for the design of immersive audio-visual environments. In *Euro-VR Eve 2010, Orsay, France*.
- Rébillat, M., Hennequin, R., Corteel, E., & Katz, B. F. (2011c). Identification of cascade of Hammerstein models for the description of nonlinearities in vibrating devices. *Journal of Sound and Vibration*, 330 :1018–1038.
- Rébillat, M., Hennequin, R., Corteel, E., & Katz, B. F. G. (2010b). Prediction of harmonic distortion generated by electro-dynamic loudspeakers using cascade of Hammerstein models. *128th Convention of the Audio Engineering Society*.
- Rébillat, M., Katz, B. F. G., & Corteel, E. (2009). SMART-I² : “Spatial Multi-user Audio-visual Real-Time Interactive Interface”, a Broadcast Application Context. *Proceedings of the IEEE 3D-TV conference*.
- Reed, M. J. & Hawksford, M. O. J. (1996). Identification of discrete Volterra series using maximum length sequences. *IEE Proceedings-circuits Devices and Systems*, 143(5) :241–248.
- Ribeiro, P. & Petyt, M. (2000). Non-linear free vibration of isotropic plates with internal resonance. *International Journal of Non-Linear Mechanics*, 35(2) :263–278.
- Riecke, B., Schulte-Pelkum, J., Caniard, F., & Balthoff, H. (2005). Influence of auditory cues on the visually-induced self-motion illusion (circular vection) in virtual reality.
- Rietveld, T. & van Hout, R. (2005). *Statistics in Language Research : Analysis of Variance*. Mouton de Gruyter, Berlin - New York.
- Rodet, X., Lambert, J., Cahen, R., Gaudy, T., Guedy, F., Gosselin, F., & Mobuchon, P. (2005). Study of haptic and visual interaction for sound and music control in the phase project. In *Proceedings of the 2005 conference on New interfaces for musical expression*, pages 109–114. National University of Singapore.
- Roy, R. & Kailath, T. (1989). Esprit - estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics Speech and Signal Processing*, 37(7) :984–995.

- Russell, M. K. & Schneider, A. L. (2006). Sound source perception in a two-dimensional setting : Comparison of action and nonaction-based response tasks. *Ecological Psychology*, 18(3) :223–237.
- Saito, T., Parbery, R. D., Okuno, S., & Kawano, S. (1997). Parameter identification for aluminum honeycomb sandwich panels based on orthotropic timoshenko beam theory. *Journal of Sound and Vibration*, 208(2) :271–287.
- Sanchez-Vives, M. V. & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4) :332–339.
- Sanson, J., Corteel, E., & Warusfel, O. (2008). Objective and subjective analysis of localisation accuracy in wave field synthesis. *124th Convention of the Audio Engineering Society*.
- Sanson, J. & Warusfel, O. (2010). évaluation perceptive des apports de la reproduction sonore par wfs dans une situation de concert. In *10ème Congrès Français d'Acoustique*.
- Saupe, K., Koelsch, S., & Rubsamen, R. (2010). Spatial selective attention in a complex auditory environment such as polyphonic music. *Journal of the Acoustical Society of America*, 127(1) :472–480.
- Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence : Factor analytic insights. *Presence-Teleoperators and Virtual Environments*, 10(3) :266–281.
- Schuemie, M. J., Van der Straaten, P., Krijn, M., & Van der Mast, C. A. P. G. (2001). Research on presence in virtual reality : A survey. *Cyberpsychology & Behavior*, 4(2) :183–201.
- Schwarz, D., Beller, G., Verbrugghe, B., Britton, S., et al. (2006). Real-time corpus-based concatenative synthesis with catart. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx), Montreal, Canada*, pages 279–282. Citeseer.
- Sestieri, A. & Ibrahim, S. R. (1994). Analysis of errors and approximations in the use of modal coordinates. *Journal of Sound and Vibration*, 177(2) :145–157.
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14(1) :147–152.
- Shi, Y. M., Sol, H., & Hua, H. X. (2006). Material parameter identification of sandwich beams by an inverse method. *Journal of Sound and Vibration*, 290(3-5) :1234–1255.
- Slater, M. (2003). A note on presence terminology. *PRESENCE-Connect [On-line]*.
- Slater, M. & Usoh, M. (1994). Depth of presence in virtual environments.
- Slater, M. & Wilbur, S. (1997). A framework for immersive virtual environments (five) : Speculations on the role of presence in virtual environments. *Presence-Teleoperators and Virtual Environments*, 6(6) :603–616.
- Soares, C. M. M., Defreitas, M. M., Araujo, A. L., & Pederson, P. (1993). Identification of material properties of composite plate specimens. *Composite Structures*, 25(1-4) :277–285.

- Sorokin, S. (2004). Analysis of wave propagation in sandwich plates with and without heavy fluid loading. *Journal of Sound and Vibration*, 271(3-5) :1039–1062.
- Sorokin, S. V. (2002). Analysis of vibrations and energy flows in sandwich plates bearing concentrated masses and spring-like inclusions in heavy fluid-loading conditions. *Journal of Sound and Vibration*, 253(2) :485–505.
- Sorokin, S. V. & Peake, N. (2006). On symmetry-breaking effects in propagation of waves in sandwich plates with and without heavy fluid loading. *Journal of Sound and Vibration*, 295(1-2) :114–128.
- Soto, R. M., López, M. R., Blanc, D. B., & Gazzo, M. (2008). Absolute threshold of coherence of position perception between auditory and visual sources for dialog. *125th Convention of the Audio Engineering Society*.
- Speigle, J. M. & Loomis, J. M. (1993). Auditory distance perception by translating observers. *Proceedings of IEEE Symposium on Research Frontiers in Virtual Reality, San Jose, CA, October 25-26*.
- Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, 28 (2) :61–70.
- Sporer, T. & Klehs, B. (2004). Wave field synthesis in the real world : Part 2 - in the movie theatre. *116th Convention of the Audio Engineering Society*.
- Springer, J. R., Sladeczek, C., Scheffler, M., Hochstrate, J., Melchior, F., & Frohlich, B. (2006). Combining wave field synthesis and multi-viewer stereo displays. *IEEE Virtual Reality 2006, Proceedings*, pages 237–+.
- Stan, G. B., Embrechts, J. J., & Archambeau, D. (2002). Comparison of different impulse response measurement techniques. *Journal of the Audio Engineering Society*, 50(4) :249–262.
- Stern, R. M. & Shear, G. D. (1996). Lateralization and detection of low-frequency binaural stimuli : Effects of distribution of internal delay. *Journal of the Acoustical Society of America*, 100(4) :2278–2288.
- Storms, R. L. (1998). *Auditory-Visual Cross-Modal Perception Phenomena*. PhD thesis, Naval Potgraduate School, Monterey, California.
- Storms, R. L. & Zyda, M. J. (2000). Interactions in perceived quality of auditory-visual displays. *Presence-Teleoperators and Virtual Environments*, 9(6) :557–580.
- Sugita, Y. & Suzuki, Y. (2003). Audiovisual perception - implicit estimation of sound-arrival time. *Nature*, 421(6926) :911–911.
- Tajadura-Jimenez, A. & Vastfjall, D. (2008). Auditory-induced emotion : A neglected channel for communication in human-computer interaction. *Affect and Emotion In Human-computer Interaction : From Theory To Applications*, 4868 :63–74.
- Talbot, J. P. & Woodhouse, J. (1997). The vibration damping of laminated plates. *Composites Part A-Applied Science and Manufacturing*, 28(12) :1007–1012.

- Thompson, J., Kuchera-Morin, J., Novak, M., Overholt, D., Putnam, L., Wakefield, G., & Smith, W. (2009). The allobrain : An interactive, stereographic, 3d audio, immersive virtual world. *International Journal of Human-computer Studies*, 67(11) :934–946.
- Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S. H., Loomis, J. M., & Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments ? *Presence-Teleoperators and Virtual Environments*, 13(5) :560–571.
- Tisseau, J. (2004). Réalité virtuelle et complexité - manifeste scientifique du cerv. *Rapport de l'ENIB*.
- Torras-Rosell, A. & Jacobsen, F. (2011). A new interpretation of distortion artifacts in sweep measurements. *Journal of the Audio Engineering Society*, 59(5) :283–289.
- Torvik, P. J. & Runyon, B. (2007). Modifications to the method of modal strain energy for improved estimates of loss factors for damped structures. *Shock and Vibration*, 14(5) :339–353.
- Trujillo-Ortiz, A., R., H.-W., & Trujillo-Perez, R. . (2004). RMAOV2 : Two-way repeated-measures ANOVA. A MATLAB file. url : <http://www.mathworks.com/matlabcentral/fileexchange/loadfile.do?objectId=5578>.
- Tsingos, N. & Warusfel, O. (2006). *Traité de la réalité virtuelle*, volume 4, chapter Dispositifs et interfaces de restitution sonore spatiale. Presses de l'École des Mines de Paris, Laval, France, 3 edition.
- Tsuchiya, T., Kagawa, Y., Doi, M., & Tsuji, T. (2003). Finite element simulation of non-linear acoustic generation in a horn loudspeaker. *Journal of Sound and Vibration*, 266(5) :993–1008.
- Ungar, E. E. & Kerwin, E. M. (1962). Loss factors of viscoelastic systems in terms of energy concepts. *Journal of the Acoustical Society of America*, 34(7) :954–&.
- Valbom, L. & Marcos, A. (2005). Wave : Sound and music in an immersive environment. *Computers & Graphics*, 29(6) :871 – 881.
- Valjamae, A. (2005). *Self-motion and Presence in the Perceptual Optimization of a Multisensory Virtual Reality Environment*. PhD thesis, Department of Signals and System, Division of Communication Systems, Chalmers University of Technology.
- Vastfjall, D. (2003). The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. *Cyberpsychology & Behavior*, 6(2) :181–188.
- Vastfjall, D., Larsson, P., & Kleiner, M. (2002). Emotion and auditory virtual environments : Affect-based judgments of music reproduced with virtual reverberation times. *Cyberpsychology & Behavior*, 5(1) :19–32.
- Verheijen, E. (1998). *Sound reproduction by wave field synthesis*. PhD thesis, Delft University of Technology.
- Verpeaux, P., Charras, T., & Millard, A. (1988). Castem 2000 : une approche moderne du calcul des structures. *Calcul des structures et intelligence artificielle*, pages 261–271.

- Vertegaal, R. & Eaglestone, B. (1998). Looking for sound ? : selling perceptual space in hierarchically nested boxes. In *CHI 98 conference summary on Human factors in computing systems*, pages 295–296. ACM.
- Voishvillo, A. (2006). Assessment of nonlinearity in transducers and sound systems - from thd to perceptual models. *121th Convention of the Audio Engineering Society*.
- Voishvillo, A., Terekhov, A., Czerwinski, E., & Alexandrov, S. (2004). Graphing, interpretation, and comparison of results of loudspeaker nonlinear distortion measurements. *Journal of the Audio Engineering Society*, 52(4) :332–357.
- Vroomen, J. & Keetels, M. (2006). The spatial constraint in intersensory pairing : No role in temporal ventriloquism. *Journal of Experimental Psychology-Human Perception and Performance*, 32(4) :1063–1071.
- Walker, G., Bowskill, J., Hollier, M., & McGrath, A. (2000). Telepresence : Understanding people as content. *Presence-Teleoperators and Virtual Environments*, 9(2) :119–136.
- Wallace, C. E. (1972). Radiation-resistance of a rectangular panel. *Journal of the Acoustical Society of America*, 51(3) :946–&.
- Wallace, C. E. (1987). The acoustic radiation damping of the modes of a rectangular panel. *Journal of the Acoustical Society of America*, 81(6) :1787–1794.
- Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, 158(2) :252–258.
- Wanderley, M. & Depalle, P. (2004). Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4) :632 – 644.
- Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth-perception in virtual environments. *Vision Research*, 35(19) :2731–2736.
- Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, 5(10) :834–862.
- Weaver, R. L. (1989). Spectral statistics in elastodynamics. *Journal of the Acoustical Society of America*, 85(3) :1005–1013.
- Wiest, W. M. & Bell, B. (1985). Stevens exponent for psychophysical scaling of perceived, remembered, and inferred distance. *Psychological Bulletin*, 98(3) :457–470.
- Wightman, F. L. & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *Journal of the Acoustical Society of America*, 91(3) :1648–1661.
- Witmer, B. G. & Singer, M. J. (1998). Measuring presence in virtual environments : A presence questionnaire. *Presence-Teleoperators and Virtual Environments*, 7(3) :225–240.

- Wittek, H., Kerber, S., Rumsey, F., & Theile, G. (2004). Spatial perception in wave field synthesis rendered sound fields : Distance of real and virtual nearby sources. In *Audio Engineering Society Convention 116*.
- Woszczyk, W., Bech, S., & Hansen, V. (1995). Interactions between audio-visual factors in a home theater system : Definition of subjective attributes. *99th Convention of the Audio Engineering Society*.
- You, J., Reiter, U., Hannuksela, M. M., Gabbouj, M., & Perkis, A. (2010). Perceptual-based quality assessment for audio-visual services : A survey. *Image Communication*, 25 :482–501.
- Yu, S. D. & Cleghorn, W. L. (2005). Free flexural vibration analysis of symmetric honeycomb panels. *Journal of Sound and Vibration*, 284(1-2) :189–204.
- Zahorik, P. (2001). Estimating sound source distance with and without vision. *Optometry and Vision Science*, 78(5) :270–275.
- Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics. *Journal of the Acoustical Society of America*, 111(4) :1832–1846.
- Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans : A summary of past and present research. *Acta Acustica United with Acustica*, 91(3) :409–420.